

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2001, 10:44:24 ; Search time 379.17 Seconds
(without alignments)
8147.469 Million cell updates/sec

Title: US-09-269-874-1

Perfect score: 4920

Sequence: 1 atgaagatcatatctttt.....tattatcacgttcaatttaa 4920

Scoring table:

IDENTITY-MUC
Gapop 10.0, Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_0601:*

- 1: /SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT:*
- 2: /SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT:*
- 3: /SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT:*
- 4: /SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT:*
- 5: /SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT:*
- 6: /SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT:*
- 7: /SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT:*
- 8: /SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT:*
- 9: /SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT:*
- 10: /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT:*
- 11: /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT:*
- 12: /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT:*
- 13: /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT:*
- 14: /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT:*
- 15: /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT:*
- 16: /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT:*
- 17: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT:*
- 18: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT:*
- 19: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:*
- 20: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:*
- 21: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:*
- 22: /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4900.8	99.6	4940	19	AAV35363
2	4779.2	97.1	5760	6	AAV50530
3	2812.8	57.2	4940	19	AAV21451
4	2694.8	54.8	5181	16	AAO80911
5	1798.8	36.6	1950	18	AAV97956
6	1785.4	36.3	1897	18	AAV97958
7	1741.2	35.4	1896	18	AAV97957
8	1063.4	21.6	1088	20	AAV56009
9	1063.4	21.6	1088	20	AAV56009
10	641.4	13.0	1077	22	AAV56009
11	601	12.2	1065	20	AAV56009

12	601	12.2	1065	20	AAV5586	Merozoite surface
13	588	12.0	1140	20	AAV56021	Merozoite surface
14	588	12.0	1140	20	AAV56021	Merozoite surface
15	535	10.9	1068	9	AAV81151	DNA sequence encod
16	525.8	10.7	660	9	AAV81148	DNA sequence encod
17	448.2	9.1	737	6	AAV50355	Storage-specific.
18	439.4	8.9	786	22	AAV68978	Merozoite surface
19	435.2	8.8	668	9	AAV82176	31-1 Repeated Dete
20	404.2	8.2	462	9	AAV81149	DNA sequence encod
21	402.6	8.2	456	9	AAV81150	DNA sequence encod
22	282.6	5.7	936	22	AAV58252	Oligonucleotide D1
23	282.6	5.7	936	22	AAV58254	Oligonucleotide D1
24	282.6	5.7	936	22	AAV58257	Oligonucleotide D1
25	282.6	5.7	936	22	AAV58259	Oligonucleotide D2
26	282.6	5.7	936	22	AAV58262	Oligonucleotide D2
27	282.6	5.7	936	22	AAV58265	Oligonucleotide D2
28	280.4	5.7	936	22	AAV58252	Oligonucleotide D1
29	280.4	5.7	936	22	AAV58254	Oligonucleotide D1
30	280.4	5.7	936	22	AAV58257	Oligonucleotide D1
31	280.4	5.7	936	22	AAV58259	Oligonucleotide D2
32	280.4	5.7	936	22	AAV58262	Oligonucleotide D2
33	280.4	5.7	936	22	AAV58265	Oligonucleotide D2
34	243.4	4.9	5940	21	AAV70105	Plasmodium falci
35	240.2	4.9	306	6	AAV50354	Storage-specific.
36	232.8	4.7	354	18	AAV80403	PfMSP1(p19)A codin
37	232.8	4.7	354	18	AAV80403	PfMSP1(p19)A codin
38	207.2	4.2	333	22	AAV68977	Merozoite surface
39	188.4	3.8	11922	21	AAV70187	Plasmodium falci
40	185.4	3.8	387	18	AAV80404	PfMSP1(p19)S codin
41	185.4	3.8	387	18	AAV80404	PfMSP1(p19)S codin
42	181.8	3.7	3579	21	AAV70099	Plasmodium falci
43	178	3.6	5409	21	AAV70151	Plasmodium falci
44	176.8	3.6	6033	21	AAV70152	Plasmodium falci
45	176.4	3.6	7458	21	AAV70106	Plasmodium falci

ALIGNMENTS

RESULT 1	
ID AAV35363	standard; DNA; 4940 BP.
AC AAV35363;	
XX	
DT 23-SEP-1998	(first entry)
DE	
XX	
KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;	
KW monoclonal antibody; passive immunisation; parasite; ss.	
XX	
OS Plasmodium falciparum.	
XX	
FH Key	Location/Qualifiers
FT CDS	10..4929
FT	/tag= a
FT	/product= gp190
XX	
PN	
MO9814583-A2.	
PD	
XX	
PF 02-OCT-1997;	97WO-EP05441.
XX	
PR 02-OCT-1996;	96DE-4040817.
XX	
PA (BUA/)	BUJARD H.
XX	
PI Bujard H, Pan W, Tolle R;	
XX	
DR WPI: 1998-240088/21.	
DR P-PSDB; AAV54145.	

XX Recombinant production of complete gp190/MSP-1 Plasmodium surface
PT protein - useful in anti-malaria vaccines, also stabilising genes by
PT reducing their AT content
XX

PS Example 1: Fig 3c; 48pp; German.

XX This sequence encodes the Plasmodium falciparum gp190/MSP-1 (merozoite
CC surface protein). This gene is used in a method for stabilising the
CC gene sequences by reducing the AT content. Such products are useful in
CC vaccines against malaria or for producing monoclonal antibodies (for
CC passive immunisation). The complete gp190 protein can now be produced
CC outside the parasite and has, at least over extended regions, the native
CC pattern of folding. Larger amounts of the protein can be produced
CC recombinantly than would be possible using the parasites as source.

SO Sequence 4940 BP; 2196 A; 597 C; 687 G; 1460 T; 0 other:

Query Match 99.6%; Score 4900.8; DB 19; Length 4940;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 4908; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 atgaagatcatattctttagtgcattctttttttattataaacacaaatgtga 60
DB 10 atgaagatcatattctttagtgcattctttttttattataaacacaaatgtga 69
QY 61 accatataagttatcaagaactgtcaaaaaactagaacattagaagatgcagtttg 120
DB 70 acacatagaagttatcaagaactgtcaaaaaactagaacattagaagatgcagtttg 129
QY 121 acaggtatagttatcttcaaaaaagaaatgttatataatgaagaacagttgagaca 180
DB 130 acaggtatagttatcttcaaaaaagaaatgttatataatgaagaacagttgagaca 189
QY 181 gctgttcaactagttacactgtgttcaaaaggttcagttgttcaggttcaggtg 240
DB 190 gctgttcaactagttacactgtgttcaaaaggttcagttgttcaggttcaggtg 249
QY 241 tcaagttctcaggttgctcaggttgctcaggttgctcaggttgctcaggttgctc 300
DB 250 tcaagttctcaggttgctcaggttgctcaggttgctcaggttgctcaggttgctc 309
QY 301 gcttcaagtttgctcaggttgctcaggttgctcaggttgctcaggttgctcaggt 360
DB 310 gcttcaagtttgctcaggttgctcaggttgctcaggttgctcaggttgctcaggt 369
QY 361 gatgttaaatcttaacgtgtatataaacacagagttacgaataattactgttaactata 420
DB 370 gatgttaaatcttaacgtgtatataaacacagagttacgaataattactgttaactata 429
QY 421 gaactcaaatctcctcaactccttgatttaactaatacatatgttaactgttgatcaat 480
DB 430 gaactcaaatctcctcaactccttgatttaactaatacatatgttaactgttgatcaat 489
QY 481 atcatagttttcaaaaataatgaatgagatataagaagaaatgaagaattatataaa 540
DB 490 atcatagttttcaaaaataatgaatgagatataagaagaaatgaagaattatataaa 549
QY 541 ttaaaacttttatttgatttaataagaagaaatgaatgagatataagaagaaatgaaga 600
DB 550 ttaaaacttttatttgatttaataagaagaaatgaatgagatataagaagaaatgaaga 609
QY 601 tgcataatccttcaaatcttaaaatcgtgcaaatgaaatagaagttactataaaact 660
DB 610 tgcataatccttcaaatcttaaaatcgtgcaaatgaaatagaagttactataaaact 669
QY 661 gtttgcgatatagaagaacattagacaatattaaagatataggaagaatggaagat 720
DB 670 gtttgcgatatagaagaacattagacaatattaaagatataggaagaatggaagat 729
QY 721 tacattaaaaaaaataaaaaaacattagaataatataatgaatgaatgaagaagtaag 780

DB 730 tacattaaaaaaaataaaaaaacattagaataatataatgaatgaatgaagaagtaag 789
QY 781 aaaaactattataaaaataagaatgcaacttaagaagaataaaaaaataataaccaa 840
DB 790 aaaaactattataaaaataagaatgcaacttaagaagaataaaaaaataataaccaa 849
QY 841 gctcaatagatcttcttatttcaataaacaattagaagaacacataatlaataagc 900
DB 850 gctcaatagatcttcttatttcaataaacaattagaagaacacataatlaataagc 909
QY 901 gttttagaaaacgatttgcaactttaaaaaaataagaagaacattagaagattactgat 960
DB 910 gttttagaaaacgatttgcaactttaaaaaaataagaagaacattagaagattactgat 969
QY 961 aagataataaatttaaaatcccccacggcaattctggaatcaccaatctctc 1020
DB 970 aagataataaatttaaaatcccccacggcaattctggaatcaccaatctctc 1029
QY 1021 cttagataagacaataaaatcgaagacagaaagaaataagaagattgccaact 1080
DB 1030 cttagataagacaataaaatcgaagacagaaagaaataagaagattgccaact 1089
QY 1081 attaaattacattgtagttatttactgacacttgaaattagaatactatlaaga 1140
DB 1090 attaaattacattgtagttatttactgacacttgaaattagaatactatlaaga 1149
QY 1141 gaaaaataataaatttgattataagtgcaaaagttgaagaagaatgaatacctgaacc 1200
DB 1150 gaaaaataataaatttgattataagtgcaaaagttgaagaagaatgaatacctgaacc 1209
QY 1201 aatgaatcccaatggagttacttactcttctgatalaagacataaactgcttta 1260
DB 1210 aatgaatcccaatggagttacttactcttctgatalaagacataaactgcttta 1269
QY 1261 aatgaatcccaatggagttacttactcttctgatalaagacataaactgcttta 1320
DB 1270 aatgaatcccaatggagttacttactcttctgatalaagacataaactgcttta 1329
QY 1321 aaaaactatatactgataatgaagaagaataatctataatgaataaaggaaaaaatt 1380
DB 1330 aaaaactatatactgataatgaagaagaataatctataatgaataaaggaaaaaatt 1389
QY 1381 aaaaactatatactgataatgaagaagaataatctataatgaataaaggaaaaaatt 1440
DB 1390 aaaaactatatactgataatgaagaagaataatctataatgaataaaggaaaaaatt 1449
QY 1441 ttaaatgataatacaaaaagataatgaagaataatctataatgaataaaggaaaaaatt 1500
DB 1450 ttaaatgataatacaaaaagataatgaagaataatctataatgaataaaggaaaaaatt 1509
QY 1501 aataataataatgaatttaacttaatttcgaaaaaatgagtggaagaatataataaa 1560
DB 1510 aataataataatgaatttaacttaatttcgaaaaaatgagtggaagaatataataaa 1569
QY 1561 gttgagaataattacacacataaacttctgcaatcctatgaagaatcttaacaatact 1620
DB 1570 gttgagaataattacacacataaacttctgcaatcctatgaagaatcttaacaatact 1629
QY 1621 gaaaagttacaagaagctcttaataatataagaagattatctttaaaggaatagtagtt 1680
DB 1630 gaaaagttacaagaagctcttaataatataagaagattatctttaaaggaatagtagtt 1689
QY 1681 gaaaagttacaagaagctcttaataatataagaagattatctttaaaggaatagtagtt 1740
DB 1690 gaaaagttacaagaagctcttaataatataagaagattatctttaaaggaatagtagtt 1749
QY 1741 tttagttgaataatttaaaaaagataagaagaagcttttgaaaaaaataattactaaagc 1800
DB 1750 tttagttgaataatttaaaaaagataagaagaagcttttgaaaaaaataattactaaagc 1809
QY 1801 gaaaataaacagatgaaaaaattttagaagttactgacattgttaaaaglacaaagtcca 1860
DB 1810 gaaaataaacagatgaaaaaattttagaagttactgacattgttaaaaglacaaagtcca 1869

QY	1861	aaagtltattaa	tgacaacaaat	tgagaaat	ttaaaaa	agactcaat	tgattttaaaaat	19270
Db	1870	aaagttttattaa	tgaaacaaat	ttgcagat	taaaaaa	gactccaat	tgattttaaaaat	19292
QY	1921	gtagatttaaaa	gcataata	tgcttcccaat	ctcttaaaa	caagaanaat	taagaagaa	19680
Db	1930	gtagaattaaaa	caataa	taacat	gtgtcccaat	ctcttaaaa	caagaanaat	19699
QY	1981	collatttaatt	atgtgttga	aaaaa	agaanaat	tgatnaat	ttaaagtgttcag	20408
Db	1990	ccttattattaa	atttggtt	gaaaaaaa	agatttgat	naaat	ttaaagtgttcag	20493
QY	2041	gtagaatcat	tgataaa	tgaaagaaaaa	aaacataaaa	caagaag	gtccaat	21000
Db	2050	gtagatcat	tgataaa	tgaaagaaaaa	aaacataaaa	caagaag	gtccaat	21009
QY	2101	tcggaaccat	taaacccga	aggaat	taaacagca	agcaat	cttaaaaaa	21600
Db	2110	tcggaaccat	taaacccga	aggaat	taaacagca	agcaat	cttaaaaaa	21699
QY	2161	gcagatctgt	tttaga	aggaaga	tttaag	taacagccca	gacaaaga	22200
Db	2170	gcagatctgt	tttaga	aggaaga	tttaag	taacagccca	gacaaaga	22299
QY	2221	caacaccag	taacag	taaccag	taacagaa	gcaaaagccaa	gtcccaacccca	22800
Db	2230	caacaccag	taacag	taaccag	taacagaa	gcaaaagccaa	gtcccaacccca	22899
QY	2281	ccagtaaa	taataaac	ctgaaaa	tglttcccaa	ttagat	taacttgaaaa	23400
Db	2290	ccagtaaa	taataaac	ctgaaaa	tglttcccaa	ttagat	taacttgaaaa	23499
QY	2341	tttttaaat	ctctat	atata	atgtacga	aatat	ttgttttcaact	24000
Db	2350	tttttaaat	ctctat	atata	atgtacga	aatat	ttgttttcaact	24099
QY	2401	gaaagat	ataataa	caata	taaat	taacaag	ggagaa	24600
Db	2410	gaaagat	ataataa	caata	taaat	taacaag	ggagaa	24699
QY	2461	gaccat	tagagat	atgtt	taata	talaca	aaataac	25200
Db	2470	gaccat	tagagat	atgtt	taata	talaca	aaataac	25299
QY	2521	gatagct	taaaacat	agttt	taacacacat	ta	tgaaatt	25800
Db	2530	gatagct	taaaacat	agttt	taacacacat	ta	tgaaatt	25899
QY	2581	tgtaatt	ataataa	actta	ggata	tgacaaat	ttat	26400
Db	2590	tgtaatt	ataataa	actta	ggata	tgacaaat	ttat	26499
QY	2641	aaag	latccacat	ctgtg	taaaaac	cttccaag	ttcaacat	27000
Db	2650	aaag	latccacat	ctgtg	taaaaac	cttccaag	ttcaacat	27099
QY	2701	cctcaag	ataaa	cccgga	agtaag	tgcaaa	tgat	27600
Db	2710	cctcaag	ataaa	cccgga	agtaag	tgcaaa	tgat	27699
QY	2761	aaatgt	taaaat	atttgga	aaacata	gtagctt	tgga	28200
Db	2770	aaatgt	taaaat	atttgga	aaacata	gtagctt	tgga	28299
QY	2821	gaat	taataag	tc	aaa	aaag	taag	28800
Db	2830	gaat	taataag	tc	aaa	aaag	taag	28899
QY	2881	acatt	ttataa	tg	aat	cttttaca	aat	29400
Db	2890	acatt	ttataa	tg	aat	cttttaca	aat	29499

QY	2941	ttgaatgatacaaaaaaggagaatctagaagaatataataataaaaaaac	3000
Db	2950	ttgaatgatacaaaaaaggagaatctagaagaatataataataaaaaaac	3009
QY	3001	ttacagttatcaattgattatataataataataataataataagaattttat	3066
Db	3010	ttacagttatcaattgattatataataataataataataataataagaattttat	3065
QY	3061	aaaaagaacaacgttggtaaatataaatccaattcaaaaactctctattaaaaa	3120
Db	3070	aaaaagaacaacgttggtaaatataataatcaattcaaaaactctctattaaaaa	3129
QY	3121	caattagataccaataatgaaatccaattaaaccaagaatgattatacaacaattctc	3180
Db	3130	caattagataccaataatgaaatccaattaaaccaagaatgattatacaacaattctc	3189
QY	3181	gtttctttaacaaaaaaaagaagctggaatctgcaaacctggaacaccttagaaaac	3240
Db	3190	gtttctttaacaaaaaaaagaagctggaatctgcaaaaactgaaaacaccttagaaaac	3249
QY	3241	acaaaatattatctggaacattataaaggcttggttaaatatataatggtgaacatc	3300
Db	3250	acaaaatattatctggaacattataaaggcttggttaaatatataatggtgaacatc	3309
QY	3301	ccattaaaaactttaagtgaagaatcaattcaaacagaagaataatattgccagtttaga	3360
Db	3310	ccattaaaaactttaagtgaagaatcaattcaaacagaagaataatattgccagtttaga	3369
QY	3361	aactttaagatataagtaaatatagaaggaaaattaaagataatttaatttagaaaag	3420
Db	3370	aactttaagatataagtaaatatagaaggaaaattaaagataatttaatttagaaaag	3429
QY	3421	aaaaattatcaacttatacgaaggatatacatcattattgctggaattaaagaagta	3480
Db	3430	aaaaattatcaacttatacgaaggatatacatcattattgctggaattaaagaagta	3489
QY	3481	ataaaaaaataaaaaatataacaggttaattctcccaagtgaaataatacggatgttaacat	3540
Db	3490	ataaaaaaataaaaaatataacaggttaattctcccaagtgaaataatacggatgttaacat	3549
QY	3541	gcattagaactctcaaaaaaattctcccaagaaggaaacgagtgtgcaacagtgtaagt	3600
Db	3550	gcattagaactctcaaaaaaattctcccaagaaggaaacgagtgtgcaacagtgtaagt	3609
QY	3601	gaaatgtgaccgacacattaggaaacaagaacaaagaacacccgcatacaactatata	3660
Db	3610	gaaatgtgaccgacacattaggaaacaagaacaaagaacacccgcatacaactatata	3669
QY	3661	ggagcagagcttaacacacataacacacatacacaatgctgcagatgaagttaagtagcta	3720
Db	3670	ggagcagagcttaacacacataacacacatacacaatgctgcagatgaagttaagtagcta	3729
QY	3721	atcatalgatactatatttggagaatccgaagaagattatgatttagacaagtagta	3780
Db	3730	atcatalgatactatatttggagaatccgaagaagattatgatttagacaagtagta	3789
QY	3781	acaggaagaagcagaactccttcctgtaattgataaatactctcaaaatgaaaaatga	3840
Db	3790	acaggaagaagcagaactccttcctgtaattgataaatactctcaaaatgaaaaatga	3849
QY	3841	tatgaggtttatataatttaaaaactcttagcaggtgttataagaagtttaaaaaacaatca	3900
Db	3850	tatgaggtttatataatttaaaaactcttagcaggtgttataagaagtttaaaaaacaatca	3909
QY	3901	gaaaatacgttatgacattaaatgttaatgttaagttaagataatttaaatcagacttaac	3960
Db	3910	gaaaatacgttatgacattaaatgttaatgttaagttaagataatttaaatcagacttaac	3969
QY	3961	aaaagtgaaaatttcaaaaatgttttagatatacagatttatccatatagaagatttaaca	4020
Db	3970	aaaagtgaaaatttcaaaaatgttttagatatacagatttatccatatagaagatttaaca	4029
QY	4021	tcaagtaattatgtgtccaagaatccaataaattcttataataagaagaaagaatataa	4080

Db	4030	tcaagtaattatgctgccaagagtcocatataatttcttaataaagaagaaaaagatataa	4089
Oy	4081	ttcttaagcagttataatctatbaltaaagatccaatagatacagatataaatttggcaat	4140
Db	4090	tctctaagcagttataatctatbaltaaagatccaatagatacagatataaatttggcaat	4149
Oy	4141	gagctcttgatatactataaatactatcccgaaaaataaaatcgatttgattcaatt	4200
Db	4150	gagctcttgatatactataaatactatcccgaaaaataaaatcgatttgattcaatt	4209
Oy	4201	aaaaaatalatcaacgcacaacaacaggtgaaatgagaaataccctcccttlaaacat	4260
Db	4210	aaaaaatalatcaacgcacaacaacaggtgaaatgagaaataccctcccttlaaacat	4269
Oy	4261	attgagacctataataaaccagttaatgtataaattgattatcttgcattccaattagaa	4320
Db	4270	attgagacctataataaaccagttaatgtataaattgattatcttgcattccaattagaa	4329
Oy	4321	gcaaaattcttcaaatatatacatatgtggaatccaacgtagaaggttaaaataaagaactt	4380
Db	4330	gcaaaattcttcaaatatatacatatgtggaatccaacgtagaaggttaaaataaagaactt	4389
Oy	4381	aattacttaaaacaaatctccagacaacaaatttggcagattcttaaaaaaataaacaaattcgtt	4440
Db	4390	aattacttaaaacaaatctccagacaacaaatttggcagattcttaaaaaaataaacaaattcgtt	4449
Oy	4441	ggatttcgtgattatctcaaacagattataacacataaacttatgtgacaagttccctgatt	4500
Db	4450	ggatttcgtgattatctcaaacagattataacacataaacttatgtgacaagttccctgatt	4509
Oy	4501	acaagatctggtttttggaaaaatctctgcctaaacccggttttctcaatttccctggatggaaac	4560
Db	4510	acaagatctggtttttggaaaaatctctgcctaaacccggttttctcaatttccctggatggaaac	4569
Oy	4561	ttgcaagagttatgttaaacacattctcacacaaccaatgtgtlaaaaaacaaatgtgccacaat	4620
Db	4570	ttgcaagagttatgttaaacacattctcacacaaccaatgtgtlaaaaaacaaatgtgccacaat	4629
Oy	4621	tctgtagtgccttcagacattttagatgtgaagagaagaatgtgtaattgtttatataacaaa	4680
Db	4630	tctgtagtgccttcagacattttagatgtgaagagaagaatgtgtaattgtttatataacaaa	4689
Oy	4681	caagaagtgatataaattgtgttgaaaaatccaacatccactttgaaagaaaataatgtgtgaa	4740
Db	4690	caagaagtgatataaattgtgttgaaaaatccaacatccactttgaaagaaaataatgtgtgaa	4749
Oy	4741	tgtgatacagatgccaaatgttaccggaagaagaattccagtagacaacggaagaagaataccaa	4800
Db	4750	tgtgatacagatgccaaatgttaccggaagaagaattccagtagacaacggaagaagaataccaa	4809
Oy	4801	tgtgatacgtatacctaaccgtgattctctatccaatttcgatagtgatttcttcgaagtctctc	4860
Db	4810	tgtgatacgtatacctaaccgtgattctctatccaatttcgatagtgatttcttcgaagtctctc	4869
Oy	4861	aactcttaagaataccaattcttatataacccaatcagtgtaatacttatacagttcaatttaa	4920
Db	4870	aactcttaagaataccaattcttatataacccaatcagtgtaatacttatacagttcaatttaa	4929

XX	Key	Location/Qualifiers
EH	Key	216..5179
ET	CDS	/*tag= a
FT		
XX		
PN	Ep154454-A.	
XX		
PD	11-SEP-1985.	
XX		
PE	21-FEB-1985;	85EP-0301173.
XX		
PR	26-SEP-1984;	84GB-0024340.
PR	22-FEB-1984;	84GB-0004692.
PR	21-FEB-1985;	85GB-0004429.
XX		
PA	(WELL) WELLCOME FOUNDATION LTD.	
XX		
PI	Holder A, Sandhu J, Odink K, Lockyer M, Riveros-Moreno V;	
DR	WPI: 1985-224845/37.	
XX	P-PSDB; AAP50777.	
XX		
PT	Cloned DNA sequence encoding plasmodium falciparum protein -	
PT	useful for expressing the protein for use in vaccines against	
PT	malaria	
XX		
PS	Claim 2; Fig 1; 51pp; English.	

CC The sequence encoding the P195 protein of *Plasmodium falciparum*
CC (AA050530) and a peptide comprising at least one of its epitopes
CC (see AA050777) are claimed. Also claimed is a vaccine for inducing
CC immunity to malaria comprising the novel peptide or P195 or a
CC peptide comprising at least one epitope when derived from the new
CC DNA sequence, together with a carrier.
XX
SQ Sequence 5760 BP, 2565 A, 630 C, 725 G, 1840 T, 0 other:

[illegible]

QY 481 attcatggttccaattatttaattggtgatatgaagaattaaatgaattattatataa 540
|||||
Db 696 attcatggttccaattatttaattggtgatatgaagaattaaatgaattattatataa 755
QY 541 ttaaaccttatttcttatttatttaagaagcaaaattaaatgattgattgctaaat 600
|||||
Db 756 ttaaaccttatttatttatttatttaagaagcaaaattaaatgattgattgctaaat 815
QY 601 tctcaaatcccttcaacttcaaatctgtgcaaatgtaagttagacttaaaaaact 660
|||||
Db 816 tctcaaatcccttcaacttcaaatctgtgcaaatgtaagttagacttaaaaaact 875
QY 661 ggtgtcggtatagaagaacccattagaacatttaagaattgtagaagaatggaag 720
|||||
Db 876 ggtgtcggtatagaagaacccattagaacatttaagaattgtagaagaatggaag 935
QY 721 tacatttaaaaaaataaaaaacccattagaacattataatgtaattatgaagaatga 780
|||||
Db 936 tacatttaaaaaaataaaaaacccattagaacattataatgtaattatgaagaatga 995
QY 781 aaaaacattgataaaaaataagaatgcactaaagaagaagaaaaaaatataccaa 840
|||||
Db 996 aaaaacattgataaaaaataagaatgcactaaagaagaagaaaaaaatataccaa 1055
QY 841 gtttcaattatgctttcttatttacaataaaccattagaagaagcacataatttaagc 900
|||||
Db 1056 gtttcaattatgctttcttatttacaataaaccattagaagaagcacataatttaagc 1115
QY 901 gttttagaaaaacgtatgtacacctttaaanaaaaatgnaaacatlaaggaattactgt 960
|||||
Db 1116 gttttagaaaaacgtatgtacacctttaaanaaaaatgnaaacatlaaggaattactgt 1175
QY 961 aagataaatgaaattgaagaatcccccacccggtcccaattctggaatatcacccaactctc 1020
|||||
Db 1176 aagataaatgaaattgaagaatcccccacccggtcccaattctggaatatcacccaactctc 1235
QY 1021 ctgtgaagaagaanaaaatcgagagacacgaanaagaataaagaatttgccaact 1080
|||||
Db 1236 ctgtgaagaagaanaaaatcgagagacacgaanaagaataaagaatttgccaact 1295
QY 1081 attaaatttaattatgatatggttatttactgtaccactgtaatttaagataactatlaaga 1140
|||||
Db 1296 attaaatttaattatgatatggttatttactgtaccactgtaatttaagataactatlaaga 1355
QY 1141 gaaaaaataaaaaataatgtgataaagtgcaaaagttgnaacaaaggaatcaactgaacc 1200
|||||
Db 1356 gaaaaaataaaaaataatgtgataaagtgcaaaagttgnaacaaaggaatcaactgaacc 1415
QY 1201 aatgaatatccaaatggaatttacttacttctgtgataaagaatttaacaatgcttta 1260
|||||
Db 1416 aatgaatatccaaatggaatttacttacttctgtgataaagaatttaacaatgcttta 1475
QY 1261 aatgaacttaattcttctgtgatttaattcaatccattgattatatacaaaagacaaagt 1320
|||||
Db 1476 aatgaacttaattcttctgtgatttaattcaatccattgattatatacaaaagacaaagt 1535
QY 1321 aaaaacatatatactgataatgaagaanaaaatcacaatgnaattgaagaaaaaact 1380
|||||
Db 1536 aaaaacatatatactgataatgaagaanaaaatcacaatgnaattgaagaaaaaact 1595
QY 1381 aaaaatgaaaaaaaatggaattgtaattgtaaaaaatcttaagaagaacgatttaagtt 1440
|||||
Db 1596 aaaaatgaaaaaaaatggaattgtaattgtaaaaaatcttaagaagaacgatttaagtt 1655
QY 1441 ttaaatgataatacaaaagaatatagaanaaatlaacttaatgnaatttatgataagcaattc 1500
|||||
Db 1656 ttaaatgataatacaaaagaatatagaanaaatlaacttaatgnaatttatgataagcaattc 1715
QY 1501 aataataataatgatttaactaattcgaaaaaataatgattggttaagaagatacataataa 1560
|||||
Db 1716 aataataataatgatttaactaattcgaaaaaataatgattggttaagaagatacataataa 1775

QY 1561 gttgagaacttacacacataataactttgtcatccatagaaatcttaacataactt 1620
|||||
Db 1776 gttgagaacttacacacataataactttgtcatccatagaaatcttaacataactt 1835
QY 1621 gaaaagttaacaaaagcctttaaataatataatggaagattacttcttaaggaatagtagtt 1680
|||||
Db 1836 gaaaagttaacaaaagcctttaaataatataatggaagattacttcttaaggaatagtagtt 1895
QY 1681 gaaaagaattaaataataataaaaatttaataagaanaattgaanaatgaattgaaaca 1740
|||||
Db 1896 gaaaagaattaaataataataaaaatttaataagaanaattgaanaatgaattgaaaca 1955
QY 1741 ttaagttgaaaataatlaaaaagaatgagaacagcttttggaaaaaaaattactaaagc 1800
|||||
Db 1956 ttaagttgaaaataatlaaaaagaatgagaacagcttttggaaaaaaaattactaaagc 2015
QY 1801 gaaaataaacccagatgaaaaaattttagaagtaactgtaactgttaaaatgataagttcaa 1860
|||||
Db 2016 gaaaataaacccagatgaaaaaattttagaagtaactgtaactgttaaaatgataagttcaa 2075
QY 1861 aaagtttatttaattgatacaaaaattgacgaatttaaaaaagctcaattgttttaaaaat 1920
|||||
Db 2076 aaagtttatttaattgatacaaaaattgacgaatttaaaaaagctcaattgttttaaaaat 2135
QY 1921 gtagaatttaaaacataataatactgttcccaattcttaacaagaanaataagcaaga 1980
|||||
Db 2136 gtagaatttaaaacataataatactgttcccaattcttaacaagaanaataagcaaga 2195
QY 1981 ccttattatttaattgtgttgaanaaagaattgataaattaaagtgttcatgctcaag 2040
|||||
Db 2196 ccttattatttaattgtgttgaanaaagaattgataaattaaagtgttcatgctcaag 2255
QY 2041 gtagaattcattgttaataatgaagaanaaaaacataaaaacccgaagttcaatccgtaatt 2100
|||||
Db 2256 gtagaattcattgttaataatgaagaanaaaaacataaaaacccgaagttcaatccgtaatt 2315
QY 2101 tcggaacccatcaacccgaagaagaataatacaggaacagcaacttaacaacccgtgacacaa 2160
|||||
Db 2316 tcggaacccatcaacccgaagaagaataatacaggaacagcaacttaacaacccgtgacacaa 2375
QY 2161 gcaggaatctgtctttagaagaagagattcagtaacaagcaagcaagaacaaaacaaga 2220
|||||
Db 2376 gcaggaatctgtctttagaagaagagattcagtaacaagcaagcaagaacaaaacaaga 2435
QY 2221 caaccacccagttaccagttaccagttaccagttaccagttaccagttaccagttaccagtt 2280
|||||
Db 2436 caaccacccagttaccagttaccagttaccagttaccagttaccagttaccagttaccagtt 2495
QY 2281 ccagtaataataaaaactgaaatgtttccaaattgaattatcttgaanaattatataatga 2340
|||||
Db 2496 ccagtaataataaaaactgaaatgtttccaaattgaattatcttgaanaattatataatga 2555
QY 2341 tttttaaacttcatatalatgttcaacaatatatttggtttcaactcaactgaac 2400
|||||
Db 2556 tttttaaacttcatatalatgttcaacaatatatttggtttcaactcaactgaac 2615
QY 2401 gaaaagaatttaaaaacataataaatttacaagaaggaagaagaacaaatgaattgaatt 2460
|||||
Db 2616 gaaaagaatttaaaaacataataaatttacaagaaggaagaagaacaaatgaattgaatt 2675
QY 2461 gatccattagactattgttttaataatacaaaaataacatccctgaattgaattgaatt 2520
|||||
Db 2676 gatccattagactattgttttaataatacaaaaataacatccctgaattgaattgaatt 2735
QY 2521 gatagcttaacaatagttattacaaactat----- 2551
|||||
Db 2736 gatagcttaaac-atagttattacaacattattatggtttatgaaagaattggtta 2794
QY 2552 -----tatggaatttatagaagaagaatggtttgtaatttatataact 2597
|||||
Db 2795 tgaattattacttattatggaatttatagaagaagaatggtttgtaatttatataact 2854
QY 2598 taaggaataatgacaaaattaaaattatttagaggaagcgaaaaagttatccacatcgt 2657

[illegible]

Dh	3935	aatacaacatccacaataatgtcgatgtagatgagtcagcttaatcattagttccataatt	3996
Qy	3738	tggaagaaatccggaagaagattatgtagatttaggacaaagtgtacaaggaagcagttac	3797
Dh	3995	tggaagaatccgaagaagattatgtagtttaggacaagtgtacaaggaagcagttac	4056
Qy	3798	tccttcgtaattgtatacatatacttcttaaatgtgaatgtaatgtatgagtttatatt	3855
Dh	4055	tacttcgtaattgtatacatatacttcttaaatgtgaatgtaatgtatgagtttatatt	4114
Qy	3858	aaaaccttagcaggttttatagaaattttaaacaacattgaagaataatgaagtttagc	3917
Dh	4115	aaaaccttagcaggttttatagaaattttaaacaacattgaagaataatgaagtttagc	4174
Qy	3918	atttaagttaattgttagagataattttaaattcagatttaataacggtgaatttcaa	3977
Dh	4175	atttaagttaattgttagagataattttaaattcagatttaataacggtgaatttcaa	4236
Qy	3978	aaatgttttagaatccagatttaattccataataagaattacaatcaagtaattatgtgt	4037
Dh	4235	aaatgttttagaatccagatttaattccataataagaattacaatcaagtaattatgtgt	4296
Qy	4038	caaagatccataataatttctttaagaagaaaagataaattctttagaagttataa	4097
Dh	4295	caaagatccataataatttctttaagaagaaaagataaattctttagaagttataa	4356
Qy	4098	ttatattaagatccaatagataaggaataaaattgtcgaatgtgttcggatata	4157
Dh	4355	ttatattaagatccaatagataaggaataaaattgtcgaatgtgttcggatata	4414
Qy	4158	taaaatattatccgaaaataataatcagatttagatccaattaaaaaatalacaaga	4217
Dh	4415	taaaatattatccgaaaataataatcagatttagatccaattaaaaaatalacaaga	4474
Qy	4218	caaaagaagtgaanaatgagaataaccttccctttttaaacaatatgtgaacttatata	4277
Dh	4475	caaaagaagtgaanaatgagaataaccttccctttttaaacaatatgtgaacttatata	4534
Qy	4278	aacagttatataaaatgtatattgttgaattgaattcatttagaagcaaaagttctaatta	4337
Dh	4535	aacagttatataaaatgtatattgttgaattgaattcatttagaagcaaaagttctaatta	4596
Qy	4338	tacatatagagaatccaacgtagaagttlaaaataaaagacttaattacttaaaaaaat	4397
Dh	4595	tacatatagagaatccaacgtagaagttlaaaataaaagacttaattacttaaaaaaat	4654
Qy	4398	tcaagaacaaattgycgagattttiaaaaaataacaattccgttgtgaattgcgtatttc	4457
Dh	4655	tcaagaacaaattgycgagattttiaaaaaataacaattccgttgtgaattgcgtatttc	4714
Qy	4458	aacagattataacataataactatgtgaacaaagttcccttagtaacggatagttttga	4517
Dh	4715	aacagattataacataataactatgtgaacaaagttcccttagtaacggatagttttga	4774
Qy	4518	aaactgttcaaaaacggttttlatcattcattacttga-tgnaaacttgcgaagttagttaa	4576
Dh	4775	aaactgttcaaaaacggttttlatcattcattacttga-tgnaaacttgcgaagttagttaa	4834
Qy	4577	acatttcaaacaccacaaatcgcgttaaaaaaacaalg--tccacaanaatctgtgatttcaag	4634
Dh	4835	acatttcaaacaccacaaatcgcgttaaaaaaacaalg--tccacaanaatctgtgatttcaag	4894
Qy	4635	acattttagatgaagaagaagaatgttaaatgttttataaatttcaacaagaagtgagttaa	4694
Dh	4895	acattttagatgaagaagaagaatgttaaatgttttataaatttcaacaagaagtgagttaa	4954
Qy	4695	atgtgttgaanaatccaatctcactgtgtgaacgaanaataatgtgtatgtatgcagatgc	4754
Dh	4955	atgtgttgaanaatccaatctcactgtgtgaacgaanaataatgtgtatgtatgcagatgc	5014
Qy	4755	caaatgtaccgaagaagaatcaggttagcaacggaagaanaatccacatgttgaattactaa	4814
Dh	5015	caaatgtaccgaagaagaatcaggttagcaacggaagaanaatccacatgttgaattactaa	5074

Dh 1150 gagaaagataaagata tagacatctccgcaaaagtcgagcaaaaggaatcaacggaacct 1209
Qy 1201 aatgaatataccaataggaattacttctcttgctatataagatatctaagaattctta 1260
Dh 1210 aatgaatataccaataggaattacttctcttgctatataagatatctaagaattctta 1269
Qy 1261 aatgaatataccaataggaattacttctcttgctatataagatatctaagaattctta 1320
Dh 1270 aatgaatataccaataggaattacttctcttgctatataagatatctaagaattctta 1329
Qy 1321 aaaaacataatactatgaagaagaaaaatctataaataaataaataaataaataaataa 1380
Dh 1330 aagaataatactatacagacaataaggaagaattatacagaaatacagaaaggaagatc 1389
Qy 1381 aaaaatgaagaaaaaaatgaattgataataaaatctacgaagaaacagatcttaagtc 1440
Dh 1390 aaaaatgaagaaaaaaatgaattgataataaaatctacgaagaaacagatcttaagtc 1449
Qy 1441 ttaaatgataatacaaaagaaataatgaagaaatctacttaataatgataagaaatc 1500
Dh 1450 ctataagataatacctaaagataatgaagaaatctctgaacgaagataatgataatc 1509
Qy 1501 aataataataatgaatttaactaatctcgaagaaatgaatgggtaaaagataatcataa 1560
Dh 1510 aaaaataataatgactgacactgcaactcgaagaaatgataagaaacgactcttaacaa 1569
Qy 1561 gttgaaataatacacaataatgacttctgataataaataatcttaaaataatctt 1620
Dh 1570 gttgaaataatacacaataatgacttctgataataaataatcttaaaataatctt 1629
Qy 1621 gaaaatgatacaaaagctcttaataataatgaagaaatctcttaagaaataatgaatc 1680
Dh 1630 gagaagctacacaagacttaagataatgaggaactattctctggaacatctgtgtg 1689
Qy 1681 gaaaagaaatgataataataataataatgataagcaaaatgataagaaatgataagaa 1740
Dh 1690 gagaagaaatgataataataataataatgataagcaaaatgataagaaatgataagaa 1749
Qy 1741 ttaagtgaataataataataataataataataataataataataataataataataata 1800
Dh 1750 ctgtgtgagaatacttaagaagataatgaagaaatctgtgtgagaagaaatgataagaa 1809
Qy 1801 gaaaataaaacagatgaagaaatcttaagaatgataagcaaaatgataagcaaaatgata 1860
Dh 1810 gaaaataaaacagatgaagaaatcttaagaatgataagcaaaatgataagcaaaatgata 1869
Qy 1861 aagaatttaatagaacaaatgataagcaaaatgataagcaaaatgataagcaaaatgata 1920
Dh 1870 aagaatttaatagaacaaatgataagcaaaatgataagcaaaatgataagcaaaatgata 1929
Qy 1921 gtaagaatgataataataataataataataataataataataataataataataataata 1980
Dh 1930 gttgagataataataataataataataataataataataataataataataataataata 1989
Qy 1981 ccttataatgataatgataagcaaaatgataagcaaaatgataagcaaaatgataagca 2040
Dh 1990 ccaatactatacctgactcgaagaaagataagcaaaatgataagcaaaatgataagca 2049
Qy 2041 gtaagaatgataataataataataataataataataataataataataataataataata 2100
Dh 2050 gttgagataataataataataataataataataataataataataataataataataata 2109
Qy 2101 tctggaacatacaacgaagaaataataataataataataataataataataataataata 2160
Dh 2110 tctggaacatacaacgaagaaataataataataataataataataataataataataata 2169
Qy 2161 gtaagaatgataataataataataataataataataataataataataataataataata 2220
Dh 2170 gttgagataataataataataataataataataataataataataataataataataata 2229
Qy 2221 caac 2280
Dh 2230 caac 2289

Qy 2281 ccaata 2340
Dh 2290 cctgtgataataataataataataataataataataataataataataataataataataata 2349
Qy 2341 ttttaataataataataataataataataataataataataataataataataataataata 2400
Dh 2350 ttttaataataataataataataataataataataataataataataataataataataata 2409
Qy 2401 gaaaagataataataataataataataataataataataataataataataataataata 2460
Dh 2410 gaaaagataataataataataataataataataataataataataataataataataata 2469
Qy 2461 gatactgataataataataataataataataataataataataataataataataataata 2520
Dh 2470 gatactgataataataataataataataataataataataataataataataataataata 2529
Qy 2521 gatactgataataataataataataataataataataataataataataataataataata 2580
Dh 2530 gatactgataataataataataataataataataataataataataataataataataata 2589
Qy 2581 tgaataataataataataataataataataataataataataataataataataataata 2640
Dh 2590 tgaataataataataataataataataataataataataataataataataataataata 2649
Qy 2641 aagaataataataataataataataataataataataataataataataataataataata 2700
Dh 2650 aagaataataataataataataataataataataataataataataataataataataata 2709
Qy 2701 cctgaataataataataataataataataataataataataataataataataataataata 2760
Dh 2710 cctgaataataataataataataataataataataataataataataataataataataata 2769
Qy 2761 aagaataataataataataataataataataataataataataataataataataataata 2820
Dh 2770 aagaataataataataataataataataataataataataataataataataataataata 2829
Qy 2821 gaaataataataataataataataataataataataataataataataataataataata 2880
Dh 2830 gaaataataataataataataataataataataataataataataataataataataata 2889
Qy 2881 acaataataataataataataataataataataataataataataataataataataata 2940
Dh 2890 acaataataataataataataataataataataataataataataataataataataata 2949
Qy 2941 tgaataataataataataataataataataataataataataataataataataataata 3000
Dh 2950 tgaataataataataataataataataataataataataataataataataataataata 3009
Qy 3001 ttaagataataataataataataataataataataataataataataataataataataata 3060
Dh 3010 ttaagataataataataataataataataataataataataataataataataataataata 3069
Qy 3061 aaaaagataataataataataataataataataataataataataataataataataata 3120
Dh 3070 aaaaagataataataataataataataataataataataataataataataataataata 3129
Qy 3121 caata 3180
Dh 3130 caata 3189
Qy 3181 gtttcttaataataataataataataataataataataataataataataataataataata 3240
Dh 3190 gtttcttaataataataataataataataataataataataataataataataataataata 3249
Qy 3241 acaata 3300
Dh 3250 acaata 3309
Qy 3301 caata 3360
Dh 3310 caata 3369

Db	4450	ggaatgcgaaccggtctaccggtataaaccacaacaaatctctgaccgaatttctctcc	4509
Oy	4501	acaggtatggttttttgaanaactctgttaaacggttttatctcaatttacttgaaggaaac	4560
Db	4510	actgcatggtgtgttcgnaaacctccgcgaacaacagtgctgagcaatctgtcgaagcgaac	4569
Oy	4561	ttgcaagtgatgtttaaacattttacacaacccaatgcygtataaanaaacaatgtccacaat	4620
Db	4570	ctgcgaaggcatgctgaacatctctccagccaacatgcytgaagaaacagtgtcccccagat	4629
Oy	4621	tctgtagtlttcagacattttagtttgaagaagaagaatgttaattgtttttatatacaca	4680
Db	4630	agcggtctgtttcagcactctcggacgagcgcgagaagtgtcaagtgtctctcgaactacaa	4689
Oy	4681	caagaagtgataaattgtttgaanaatcccaatcttactgttaacgaaataatggtgga	4740
Db	4690	caagaagagagataaagtgctggtggaacccaacaccttactctgaatgaaanaaatggtcgg	4749
Oy	4741	tgtatgtcagatgccaatgtlaccggaagaagatctcaggtagcaacggaagaagaataca	4800
Db	4750	tgtgacgcgcatgtctaattgcaccgcggagaagaacggtcttaacggaagaatacaca	4809
Oy	4801	tgtgaatgtactaaacctgtattcttatccactttcgaatggtatttttgcagttctct	4860
Db	4810	tgcgagtgactaaagcccgactctctatccactcttcgacggatttttgtccagctct	4869
Oy	4861	aactcttgaagaatattcttcttatttaacttaactcatgtaataatatacagtttcattaa	4920
Db	4870	aattcttcggacatctcttctctcgtcgtacccaatgctgacgtcgtacagttcatctaa	4929
RESULT 4			
AAQ80911			
ID	AAQ80911 standard; cDNA; 5181 BP.		
XX	AAQ80911;		
XX	24-AUG-1995 (first entry)		
DE	Plasmodium falciptarum MSA-1 gene CDNA.		
XX			
KW	Plasmodium falciptarum MSA-1 gene; recombinant poxvirus;		
KM	multicomponent multistage malarial vaccines; immunogens;		
XX	malaria diagnosis; ss.		
OS	Plasmodium falciptarum (p486195).		
PN	W09428930-A.		
PD	22-DEC-1994.		
XX			
PF	10-JUN-1994; 94WO-US06652.		
PR	11-JUN-1993; 93US-0075783.		
PR	09-JUN-1994; 94US-0257073.		
PA	(VIR0-) VIROGENETICS CORP.		
XX			
XX	De Taisne C, Paoletti E, Tine JA;		
DR	WPI; 1995-036113/05.		
XX			
PT	Recombinant poxvirus contg. Plasmodium DNA in non-essential		
PT	region - useful in vaccines against malaria and for prodn. of		
XX	Plasmodium immunogens		
PS	Claim 3; Fig 6; 183pp; English.		
XX			
CC	AAQ80911 is the P. falciptarum MSA-1 gene cDNA sequence. New		
CC	recombinant poxviruses containing either the SBRa, ABRA, P18sp70,		
CC	AMA-1, Pf1s2j, CSP, Pf1SP2, LSA-1, LSA-1 repeatless, MSA-1,		
CC	MSA-1 (N-terminal p83 or C-terminal gp42) genes, or a combination		

Db	4450	ggaatgcgaaccggtctaccggtatlaaccacaacaactctcctgaccgaagtttctctc	4509
Oy	4501	acaggtatggttttttgaanaactctgttaaacctgtttatctcaattacttgaaggaaac	4560
Db	4510	actgcatggtgtgttcgaaaacctccgcacaacagtgctgagcaatctgtcgaagcgaac	4569
Oy	4561	ttgcaagtgatgtttaaacattttcacacaaccaatgctgataaanaaacatgtccacaat	4620
Db	4570	ctgcgaaggcatgctgaacatctctccagcaccaatgctgagaagaacagtgcccccagat	4629
Oy	4621	ctgtgatgttttcagacatttagatgtgaagaagaagatgttaattgtttatataacaaa	4680
Db	4630	agcggtctgttttcagacatcttcggacgagcgcgagaagtgcaagtgtctctcgaactacaa	4689
Oy	4681	caagaagtgataaattgtgttgaanaatccaaatctcactgttgaacgaanaatagtgtgga	4740
Db	4690	caagaagagagataaagtgctggtggaacccaacacctactcgtgaatgaaanaacatggcggg	4749
Oy	4741	tgtatgtcagatgccaatgtlaccggaagaagatctcaggtagcaacggaagaagaaatcaca	4800
Db	4750	tgtgacgcgcatgtctaattgacccggaggaagaacgagctctaacggaagaataatcaca	4809
Oy	4801	tgtgaatgtactaaacctgtattcttatccactttcgaatggtatttttgcagttctct	4860
Db	4810	tgcgagtgactaaagcccgacatctcatcactcttcgaaggtatttttgcagctct	4869
Oy	4861	aactcttagaataatcatctcttatttaacttaactcagtgtaataatataacagttcattaa	4920
Db	4870	aattcttcggacatctctcttcctcgtcatctccatcgaatgctgacgttcaagcttcatctaa	4929
RESULT 4			
AAQ80911			
ID	AAQ80911	standard; cDNA; 5181 BP.	
XX	XX	AAQ80911;	
XX	XX	24-AUG-1995 (first entry)	
XX	XX	Plasmidium falciparum MSA-1 gene CDNA.	
XX	XX	Plasmidium falciparum MSA-1 gene; recombinant poxvirus;	
KW	KW	multicomponent multistage malarial vaccines; immunogens;	
XX	XX	malaria diagnosis; ss.	
OS	OS	Plasmidium falciparum (p486195).	
XX	XX	W09428930-A.	
PN	PN	22-DEC-1994.	
XX	XX	10-JUN-1994; 94WO-US06652.	
XX	XX	11-JUN-1993; 93US-0075783.	
PR	PR	09-JUN-1994; 94US-0257073.	
XX	XX	(VIR0-) VIROGENETICS CORP.	
XX	XX	De Taisne C, Paoletti E, Tine JA;	
XX	XX	WPI; 1995-036113/05.	
XX	XX	Recombinant poxvirus contg. Plasmidium DNA in non-essential	
PT	PT	region - useful in vaccines against malaria and for prodn. of	
XX	XX	Plasmidium immunogens	
XX	XX	Claim 3; Fig 6; 183pp; English.	
XX	XX	AAQ80911 is the P. falciparum MSA-1 gene cDNA sequence. New	
CC	CC	recombinant poxviruses containing either the SBRa, ABRA, P1hsp70,	
CC	CC	AMA-1, Pf1s2j, CSP, Pf1SP2, LSA-1, LSA-1 repeatless, MSA-1,	
CC	CC	MSA-1 (N-terminal p83 or C-terminal gp42) genes, or a combination	

ID AA097956 standard; DNA; 1950 BP.
XX
AC AAT97956;
XX
DT 03-APR-1998 (first entry)
XX
DE Chimeric MSA-1 antigenic protein 1 used in a malaria vaccine.
KW Signal peptide; malaria; vaccine; merozoite surface antigen-1 peptide;
KM MSA-1; recombinant vaccinia virus; Plasmodium falciparum; anchor peptide;
KW immune response; humoral; cell-mediated; merozoite; ss.
OS Chimeric - Mammalia.
OS Chimeric - Plasmodium falciparum.
OS Chimeric - Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_feature 1..165 /tag= a
FT /note= "signal sequence added to improve
FT immunogenicity"
FT misc_feature 166..1893 /tag= b
FT /note= "Plasmodium falciparum MSA-1 peptide"
FT misc_feature 1894..1950 /tag= c
FT /note= "anchor sequence added to improve
FT immunogenicity"
FT
FN W09726911-A1.
PN
PD 31-JUL-1997.
PE 29-JAN-1997; 97WO-US01395.
PF
PR 29-JAN-1996; 96US-0593006.
PA (GEOU) UNIV GEORGETOWN.
XX
XX Davidson EA, Yang S;
PI WPI, 1997-393372/36.
DR
DX Malaria vaccine - comprises expression vector expressing fragment of
PT merozoite surface antigen
PS
PS Claim 33; Fig 2; 75pp; English.

The present sequence encodes a chimeric protein that contains a mammalian signal and anchor sequence, and a merozoite surface antigen-1 (MSA-1) carboxy terminal peptide. The C-terminal fragment of MSA-1 provides a more specific response than the complete MSA-1, and attachment of anchor and signal sequences improve the immunogenicity of the protein better than the use of an adjuvant. The chimeric protein, and expression vectors (analogues that express MSA-1 without either signal or anchor peptides), particularly in the form of recombinant vaccinia virus, are used in vaccines to prevent or treat malaria caused by Plasmodium falciparum. The vaccinia vector expresses the antigen fragment for many days, or even years, generating a long-lasting immune response (humoral and/or cell-mediated) against the merozoite form of the parasite, in humans or other animals.

Sequence 1950 BP; 830 A; 236 C; 277 G; 607 T; 0 other;

Query Match 36.6%; Score 1798.8; DE 16; Length 1950;
Best Local Similarity 95.5%; Pred. No. 3.le-197;
Matches 1862; Conservative 0; Mismatches 87; Indels 1; Gaps 1;

FH	Key	Location/Qualifiers
FT	Misc_feature	1..1839
FF		/tag= a
FT		/note= "Plasmodium falciparum MSA-1 peptide"
FT	Misc_feature	1840..1897
FF		/tag= b
FT		/note= "anchor sequence added to improve Immunogenicity"
FN		
PN	WO9726911-A1.	
XX		
PD	31-JUL-1997.	
XX		
PF	29-JAN-1997;	97WO-US01395.
XX		
PR	29-JAN-1996;	96US-0593006.
XX		
PA	(GEOU) UNITV GEORGETOWN.	
XX		
PI	Davidson EA, Yang S;	
DR	WPI; 1997-393372/36.	
XX		
PT	Malaria vaccine - comprises expression vector expressing fragment of merozoite surface antigen	
PS	Claim 33; Fig 4; 75pp; English.	
XX		
CC	The present sequence encodes a chimeric protein that contains a human anchor sequence, and a merozoite surface antigen-1 (MSA-1) carboxy terminal peptide. The C-terminal fragment of MSA-1 provides a more specific response than the complete MSA-1, and attachment of the anchor sequence improve the immunogenicity of the protein better than the use of an adjuvant. The chimeric protein, and expression vectors (analogues) that express MSA-1 without either signal or anchor peptides), particularly in the form of recombinant vaccinia virus, are used in vaccines to prevent or treat malaria caused by Plasmodium falciparum. The vaccinia vector expresses the antigen fragment for many days, or even years, generating a long-lasting immune response (humoral and/or cell-mediated) against the merozoite form of the parasite, in humans or other animals.	
SQ	Sequence 1897 BP; 815 A; 230 C; 274 G; 578 T; 0 other:	
	Query Match	36.3%; Score 1785.4; DB 18; Length 1897;
	Best Local Similarity	97.2%; Pred. No. 1e-195;
	Matches 1838; Conservative	0; Mismatches 51; Indels 2; Gaps 2.
OY	3032 ataattaaatatagaaagtatttggataaaaagaacaacgttgtaataatataatg c 3091 Db 7 acaactgaaggcttacaaagaactgtccaanaaacatagaagccttgaagatgcagtatg 66	
OY	3092 aaattaanaaaaccttactttatca-aagaacacatlagaalcaaatgtgatccactaat 3150 Db 67 acaggttatagttatttcocaaaaagaaaaaatgvttaaataagatltgatccactaat 126	
OY	3151 aaccocaagcatgtattacaacacttctgtttcttccttaacaaaaaaaaaagaagctgaa 3210 Db 127 aaccocaagcatgtattacaacacttctgtttcttccttaacaaaaaaaaaagaagctgaa 186	
OY	3211 atagcagaagaacgtgaanaacatctagaanaacacaaaaratcttgtgaacattataaga 3270 Db 187 atagcagaagaacgtgaanaacatctagaanaacacaaaaratcttgtgaacattataaga 246	
OY	3271 ctgtttaaatatcataatgttgatcatcatccattaaaacattlaagtlgaagaatcoact 3330 Db 247 ctgtttaaatatcataatgttgatcatcatccattaaaacattlaagtlgaagaatcoact 306	
OY	3331 caaacagaagaataattatgccagtttagaaaaactttaagattaaagtattagaaga 3390 Db 307 caaacagaagaataattatgccagtttagaaaaactttaagattaaagtattagaaga 366	

QY	3391	aaatuaagagtaattttaaattttagaagaagaaaaatatacactactatcaagtygatta	3430
Db	367	aaatuaagagtaattttaaattttagaagaagaaaaatatacactactatcaagtygatta	426
QY	3451	catcatttaattgtcttgataattaaagaagtaataaaaataataatatacagtaattct	3510
Db	427	catcatttaattgtcttgataattaaagaagtaataaaaataataatatacagtaattct	486
QY	3511	coaagtgaanaataaataagagttgtaacaatgcatltaagaactctacaanaaatcttccca	3570
Db	487	coaagtgaanaataaataagagttgtaacaatgcatltaagaactctacaanaaatcttccca	546
QY	3571	gaaggaacaagatgtctgcacagtttgtaagtgaagtgtgataccgcacaatttagaacaagt	3630
Db	547	gaaggaacaagatgtctgcacagtttgtaagtgaagtgtgataccgcacaatttagaacaagt	606
QY	3631	caaccaagaagaaccaagataactcctgttagggcgagttcaacaataaacaacatac	3690
Db	607	caaccaagaagaaccaagataactcctgttagggcgagttcaacaataaacaacatac	666
QY	3691	caaaatgtcgaatgtgaagtagatgaatgaatcatatgatacttattggaataccgaa	3750
Db	667	caaaatgtcgaatgtgaagtagatgaatgaatcatatgatacttattggaataccgaa	726
QY	3751	gaagattaatgtatttagagacaagtatgaacagsgaagcaactcctccgtaatt	3810
Db	727	gaagattaatgtatttagagacaagtatgaacagsgaagcaactcctccgtaatt	786
QY	3811	gatacactacttcttaaatgtgaatatgaatatgaagtttatattttaaacctttagca	3870
Db	787	gatacactacttcttaaatgtgaatatgaatatgaagtttatattttaaacctttagca	846
QY	3871	ggtgtttatagaagttttaaaaaacaattagaanaataacgttatgaatttaagttat	3930
Db	847	ggtgtttatagaagttttaaaaaacaattagaanaataacgttatgaatttaagttat	906
QY	3931	gttaagataattttaaatltaacagatttaataaacytgaanaattccaanaatgttttaga	3990
Db	907	gttaagataattttaaatltaacagatttaataaacytgaanaattccaanaatgttttaga	966
QY	3991	tcgaatttaattcataataagaatttaacatacgaatatagttgtgcaagaatccatat	4050
Db	967	tcgaatttaattcataataagaatttaacatacgaatatagttgtgcaagaatccatat	1026
QY	4051	aaatttcttaataagaanaaagagataaattcttaagcagttataatataatagaatt	4110
Db	1027	aaatttcttaataagaanaaagagataaattcttaagcagttataatataatagaatt	1086
QY	4111	tcgaatagataaggaatataaaatttctgcaatgacgttctgtgatalatatanaaatatacc	4170
Db	1087	tcgaatagataaggaatataaaatttctgcaatgacgttctgtgatalatatanaaatatacc	1146
QY	4171	gaaaaataataatcagatttagatttcaatataaanaatatataccaacgacaacaagtyga	4230
Db	1147	gaaaaataataatcagatttagatttcaatataaanaatatataccaacgacaacaagtyga	1206
QY	4231	aabgagaataacccctcccttlltaaacaaatgtgaagccttatataaagcgtttaagat	4290
Db	1207	aabgagaataacccctcccttlltaaacaaatgtgaagccttatataaagcgtttaagat	1266
QY	4291	aaagtatattatctgtcaatttcatltagaagcaaaagtcttaactatatacatatagaa	4350
Db	1267	aaagtatattatctgtcaatttcatltagaagcaaaagtcttaactatatacatatagaa	1326
QY	4351	tcaaacgttagaagttaaataaagaacttaattacttaaaaaacaatccaagaacaattg	4410
Db	1327	tcaaacgttagaagttaaataaagaacttaattacttaaaaaacaatccaagaacaattg	1386
QY	4411	gcagattttaaaaaataatacaaatctcgtttgagattgtcgatttacaacgattatbaac	4470
Db	1387	gcagattttaaaaaataatacaaatctcgtttgagattgtcgatttacaacgattatbaac	1446
QY	4471	catataactatttgaaaagctccttagtaacagatgatttttggaaaacttcttgtaa	4530

Db	1447	catataactatattgacaagaagllcctttagtaaggataygttttgaacaatctgtctaa	1506
Oy	4531	acgctttatctaatcttaacttgaatggaanaacttgcagaagratgttaacatttcaaacac	4590
Db	1507	acgctttatcttaattcttaacttgaatggaanaacttgcagaagratgttaacatttcaaacac	1566
Oy	4591	caatgctgtaaaaaaacaatgctccacaanaatlctgtatgtttcagacatttgaatgaaga	4650
Db	1567	caatgctgtaaaaaaacaatgctccacaanaattctgtagtcttcagcatttgaatgaaga	1626
Oy	4651	gaagaatgtaaatggtttatttaaatctcaacaagaagtgatataatgtgttgaataatcca	4710
Db	1627	gaagaatgtaaatggtttatttaaatctcaacaagaagtgatataatgtgttgaataatcca	1686
Oy	4711	aatctctctgttaagaanaataatgtgtgatatgtatgcagatgcgaatgtatccgagaagaa	4770
Db	1687	aatctctctgttaagaanaataatgtgtgatatgtatgcagatgcgaatgtatccgagaagaa	1746
Oy	4771	gattcagatgtaagaacggaanaagaatccaatgtgaatgaatgaacttaacccgatattctatcca	4830
Db	1747	gattcagatgtaagaacggaanaagaatccaatgtgaatgaatgaacttaacccgatattctatcca	1806
Oy	4831	cttttcagatgtaattttctgcagttcttcttaacttcttgaagaatcatcttctaat-aa	4889
Db	1807	cttttcagatgtaattttctgcagttcttcttaacttcttgaagaatcatcttcttaaat	1866
Oy	4890	actcaatgtaataattatatacagtttcaattaa	4920
Db	1867	actcaatgtaataattatatacagtttcaattaa	1897

Result	7
AAAT97957	standard; DNA; 1896 BP.
AAAT97957:	
03-APR-1998	(first entry)
Chimeric MSA-1 antigenic protein 2 used in a malaria vaccine.	
Signal peptide; malaria; vaccine; merozoite surface antigen-1 peptide; MSA-1; recombinant vaccinia virus; Plasmodium falciparum; anchor peptide; immune response; humoral; cell-mediated; merozoite; ss.	
Chimeric - Mammalia.	
Chimeric - Plasmodium falciparum.	
Key	Location/Qualifiers
misc_feature	1..165
	/*tag= a
	/note= "Signal sequence added to improve immunogenicity"
misc_feature	166..1896
	/*tag= b
	/note= "Plasmodium falciparum MSA-1 peptide"
MO9726911-A1.	
31-JUL-1997.	
29-JAN-1997:	97MO-US01395.
29-JAN-1996:	96US-0593006.
(GEOU) UNIV GEORGETOWN.	
Davidson EA, Yang S;	
WPI: 1997-393372/36.	
Malaria vaccine - comprises expression vector expressing fragment of	

PT merozoite surface antigen
XX
PS
XX Claim 33; Fig 3; 75pp; English.
XX
CC The present sequence encodes a chimeric protein that contains a mammalian
CC signal sequence, and a merozoite surface antigen-1 (MSA-1) carboxy
CC terminal peptide. The C-terminal fragment of MSA-1 provides a more
CC specific response than the complete MSA-1, and attachment of the
CC signal sequence improve the immunogenicity of the protein better than
CC the use of an adjuvant. The chimeric protein, and expression vectors
CC (analogues that express MSA-1 without either signal or anchor peptides),
CC particularly in the form of recombinant vaccinia virus, are used in
CC vaccines to prevent or treat malaria caused by *Plasmodium falciparum*. The
CC vaccinia vector expresses the antigen fragment for many days, or even
CC years, generating a long-lasting immune response (humoral and/or
CC cell-mediated) against the merozoite form of the parasite, in humans or
CC other animals.
XX
SQ Sequence 1896 BP; 813 A; 230 C; 272 G; 581 T; 0 other;

Query Match	35.4%;	Score 1741.2;	DB 18;	Length 1896;
Best Local Similarity	95.3%;	Pred. No. 1.1e-150;		
Matches 1805; Conservative	0;	Mismatches 88;	Indels 1;	Gaps 1

QY	2972	aagaagatattaaatataaaaaaaccttacagttacattgattatataataat	3031
Db	1		
QY	3032	atgaatataattgaagaattatgatataaagaacagttgataataataatgc	3091
Db	61		
QY	3092	aaattaaacttactttat-taagaacattgaattgaattgaattcacttaat	3150
Db	121		
QY	3151	aaccgaagcgtatgtatcacaaactttcgttttccttaacaaaaaaaagaagctaa	3210
Db	181		
QY	3211	atagcagaacttgaaaaacacattagaacacacaaaattattatgaacattataaaga	3270
Db	241		
QY	3271	cttgttaatatiaatgtgtgaatcatcccatiaaaacttaagtgaagatcaatt	3330
Db	301		
QY	3331	caaacagaagataattatgcagtttagaanaacttaagataaagtaattagaaga	3390
Db	361		
QY	3391	aaattaaaggaataatttaatttagaanaagaaaaattctactactcaagtgatta	3450
Db	421		
QY	3451	catcatattatgttgaattaaagaagtaataaaaaataaaattatacagttact	3510
Db	481		
QY	3511	ccaagtgaataataacggatggttaacaatgcattagaattcttacaanaaaatttccca	3570
Db	541		
QY	3571	gaaggaacagatgtgtcaacagttgttaagtgaagaatggaatccgacacattagaacaaat	3630
Db	601		
QY	3631	caaccagaacacagcatcaactcatgttagagagagttcaacacataaacacatca	3690
Db	661		


```

QY 3691 caaatgcgatacgagtagatgacgctaatcatagctacatatttggagaatccgaa 3750
    |||
Db 721 caaatgcatgatgagtagatgacgctaatcatagctacatatttggagaatccgaa 780
QY 3751 gaagattatgatttagacaagtagtaacgggagaacagtaactctctccgtaatt 3810
    |||
Db 781 gaagattatgatttagacaagtagtaacgggagaacagtaactctctccgtaatt 840
QY 3811 gatacatattcttaaatgaaatgaaatgatgaggtttatatttaaacctttagca 3870
    |||
Db 841 gatacatattcttaaatgaaatgaaatgatgaggtttatatttaaacctttagca 900
QY 3871 ggtgtttatagaaggttttaaaaaacaattagaataacgttttagacatttaagt 3930
    |||
Db 901 ggtgtttatagaaggttttaaaaaacaattagaataacgttttagacatttaagt 960
QY 3931 gtttaaggatattttaatttcaatgatttaataacgttgaataattcaaaatgttttagaa 3990
    |||
Db 961 gtttaaggatattttaatttcaatgatttaataacgttgaataattcaaaatgttttagaa 1020
QY 3991 tcagatttaattccataaagaatttaacaatgaattatgtttgcaaggtccat 4050
    |||
Db 1021 tcagatttaattccataaagaatttaacaatgaattatgtttgcaaggtccat 1080
QY 4051 aaattcttaataaagaataaagaagataaattcttaagcagttataattatagaat 4110
    |||
Db 1081 aaattcttaataaagaataaagaagataaattcttaagcagttataattatagaat 1140
QY 4111 tcataatgatacgatataaaatttgcgaatgagttcttggatataataaataattatcc 4170
    |||
Db 1141 tcataatgatacgatataaaatttgcgaatgagttcttggatataataaataattatcc 1200
QY 4171 gaaaatataaatacaagatttagattcaattaaataatacaacgaacgaagtgaa 4230
    |||
Db 1201 gaaaatataaatacaagatttagattcaattaaataatacaacgaacgaagtgaa 1260
QY 4231 aatgagaataacacctcttctttaaacaatattgagacattataaacaagttatgat 4290
    |||
Db 1261 aatgagaataacacctcttctttaaacaatattgagacattataaacaagttatgat 1320
QY 4291 aaattgattatttggatttcaattcttagaagaacgaagtttaattatacatgtgaaa 4350
    |||
Db 1321 aaattgattatttggatttcaattcttagaagaacgaagtttaattatacatgtgaaa 1380
QY 4351 tcaaacgttagaagttaaaaataaagaacttaattacttaaaaaacaattcaagaacaaatg 4410
    |||
Db 1381 tcaaacgttagaagttaaaaataaagaacttaattacttaaaaaacaattcaagaacaaatg 1440
QY 4411 gcagattttaaaaaataaacaattcogttggaattgctgattatcaacagattatcaac 4470
    |||
Db 1441 gcagattttaaaaaataaacaattcogttggaattgctgattatcaacagattatcaac 1500
QY 4471 cataataactattgcaaaagttccttagtagaggtatgttttgaaaatcctgctaaa 4530
    |||
Db 1501 cataataactattgcaaaagttccttagtagaggtatgttttgaaaatcctgctaaa 1560
QY 4531 accgttttacttaattacttgaatggaacttgcagagttgtttaaacaatttacaacac 4590
    |||
Db 1561 accgttttacttaattacttgaatggaacttgcagagttgtttaaacaatttacaacac 1620
QY 4591 caatgcgttaaaaaacaatgctccacaataatctcgatglttcgacacatttagatgaaga 4650
    |||
Db 1621 caatgcgttaaaaaacaatgctccacaataatctcgatglttcgacacatttagatgaaga 1680
QY 4651 gaagaaatgtaaatggtttatttaattcaaacgaagaaggtgataaagtgttgaanaatcca 4710
    |||
Db 1681 gaagaaatgtaaatggtttatttaattcaaacgaagaaggtgataaagtgttgaanaatcca 1740
QY 4711 aatcccaactgttaacgaagaataatgtgtgagtcgaatcccaaatctaccgaagaa 4770
    |||
Db 1741 aatcccaactgttaacgaagaataatgtgtgagtcgaatcccaaatctaccgaagaa 1800
QY 4771 gattcagtagcaaaaggaaagaataatcaatgtgaaatgtaactaaacctgattctatcca 4830

```

```

Db 1801 gattcagtagcaaaaggaaagaataatcaatgtgaaatgtaactaaacctgattctatcca 1860
QY 4831 ctlttcagtagtaatttcttcgagttcctaact 4864
Db 1861 ctlttcagtagtaatttcttcgagttcctaact 1894

RESULT 8
AA56009
ID AA56009 standard; cDNA; 1088 BP.
XX
AC AA56009;
XX
DT 31-AUG-1999 (first entry)
XX
DE Merozoite surface protein MSP-1-42 cDNA.
XX
KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal; ss.
XX
OS Plasmodium falciparum.
XX
FH Key
FT Location/Qualifiers
FT CDS 1..1086
FT /*tag= a
XX
PN MO920774-A2.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22226.
XX
PR 15-MAY-1998; 98US-0085649.
XX
PR 20-OCT-1997; 97US-0062592.
XX
PA (GEN2 ) GENZYME TRANSGENICS CORP.
XX
PI Chen LH, Meade H;
XX
DR WPI; 1999-28833/24.
DR P-PSDB; AAY09373.
XX
PT Modified malarial protein for use in anti-malarial vaccines
XX
PS Example; Fig 2; 35pp; English.
XX
CC This nucleic acid encodes a 42 kDa C-terminal portion (see AAY09373)
CC of malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The 3' end of the sequence has been
CC modified to include a 6xHis tag. The nucleic acid sequence has
CC been modified (see AA56008) according to a method of the invention
CC in order to improve expression in mammalian host cells and in
CC transgenic animals. In the modified coding sequence, 306 nucleotide
CC positions have been replaced to lower the AT content (from 76 to
CC 49.7%) and to eliminate 10 mRNA instability motifs. The encoded
CC amino acid sequence is unaltered. In another modified sequence
CC (see AA56021), a signal peptide sequence has been added and two
CC N-glycosylation sites eliminated. The invention allows expression
CC of MSP-1-42 protein in the milk of transgenic animals, and also
CC provides a DNA vaccine comprising a vector containing the altered
CC MSP-1-42 sequence.
XX
SQ Sequence 1088 BP; 454 A; 139 C; 150 G; 345 T; 0 Other:

```

Query Match 21.6%; Score 1063.4; DB 20; Length 1088;
Best local Similarity 99.9%; Pred. No. 1.4e-113;
Matches 1064; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3790 gcagtaactcttcgtaattgataacatacttctaataatgaaatgataatgaggtt 3849


```

|||||
Db      1  ggaagtaacccttcgtaattgataacacattctcaaatggaatgaaatgaatcagggct 60
Oy      3850 ttatatttaaacctttagcaggtgtttataagaatttaaaaaaacaaatagaataaac 3909
Db      61  ttatatttaaacctttagcaggtgtttataagaatttaaaaaaacaaatagaataaac 120
Oy      3910 gttatgacatttaagttaattgaagattttaaatttcacgatttaataacgtga 3969
Db      121 gttatgacatttaagttaattgaagattttaaatttcacgatttaataacgtga 180
Oy      3970 aattccaanaatggttttagaaccagatttaattccatataaagaatttaacatcaagtaat 4029
Db      181 aattccaanaatggttttagaaccagatttaattccatataaagaatttaacatcaagtaat 240
Oy      4030 tatgtgttcaagaatccatataaattcttaataaagaagaagaataattcttaagc 4089
Db      241 tatgtgttcaagaatccatataaattcttaataaagaagaagaataattcttaagc 300
Oy      4090 agttaaatattataaggaatccaatagatacagataataatttgcgaatgattctt 4149
Db      301 agttaaatattataaggaatccaatagatacagataataatttgcgaatgattctt 360
Oy      4150 ggatattataaataattatccgaaaaataataacagatttagattcaattaaaaaatat 4209
Db      361 ggatattataaataattatccgaaaaataataacagatttagattcaattaaaaaatat 420
Oy      4210 atcaacgaaacaaaggtggaataatgagaataatccttccttttaacaatattgagcc 4269
Db      421 atcaacgaaacaaaggtggaataatgagaataatccttccttttaacaatattgagcc 480
Oy      4270 ttatataaacaggttaattgataaattgatttatttgaattcaatttagaagcaaaagt 4329
Db      481 ttatataaacaggttaattgataaattgatttatttgaattcaatttagaagcaaaagt 540
Oy      4330 ctaaatatatacatatgagaatcaaacgtagaagtttaaaataaagaacttaattactta 4389
Db      541 ctaaatatatacatatgagaatcaaacgtagaagtttaaaataaagaacttaattactta 600
Oy      4390 aaaaacaattcaagacaacaattgtagaatttaaaaaaaataacaattcgttggaattgct 4449
Db      601 aaaaacaattcaagacaacaattgtagaatttaaaaaaaataacaattcgttggaattgct 660
Oy      4450 gattatcaacagattatatacattataactattatgacaagaatttccttgtagcaggtatg 4509
Db      661 gattatcaacagattatatacattataactattatgacaagaatttccttgtagcaggtatg 720
Oy      4510 gttttgaaaatctgtctataaacgltttatcttaatttacttgatggaactgtcaaggt 4569
Db      721 gttttgaaaatctgtctataaacgltttatcttaatttacttgatggaactgtcaaggt 780
Oy      4570 atgttaaacattcaacacacacaaatgctgaataaaacaatgctcacaaaattctgtgat 4629
Db      781 atgttaaacattcaacacacacaaatgctgaataaaacaatgctcacaaaattctgtgat 840
Oy      4630 ttccagacatttgaagaagaagaagaatgttaatttgaatttaataacaaagaaggt 4689
Db      841 ttccagacatttgaagaagaagaagaatgttaatttgaatttaataacaaagaaggt 900
Oy      4690 gataaatgttgaataatccaacactacttgaacgaataatgttgaatgtgtga 4749
Db      901 gataaatgttgaataatccaacactacttgaacgaataatgttgaatgtgtga 960
Oy      4750 gatgcgaatgtcaggaagaagatccaggttagcaacggaagaagaataatcactgtgaat 4809
Db      961 gatgcgaatgtcaggaagaagatccaggttagcaacggaagaagaataatcactgtgaat 1020
Oy      4810 actaaacggtattctatccacttttcagtgatttttcaggt 4854
Db      1021 actaaacggtattctatccacttttcagtgatttttcaggt 1065

```

```

AA25587
ID      AAX25587 strand: cDNA; 1088 BP.
XX
AC      AAX25587;
XX
DT      02-AUG-1999 (first entry)
XX
DE      Merozoite surface protein MSP-1-42 cDNA.
XX
KW      MSP-1; merozoite surface protein; malaria; vaccine;
KW      protein engineering; protein expression; codon usage;
KW      transgenic animal; ss.
XX
OS      Plasmodium falciparum.
XX
FH      Key
FH      CDS
FH      Location/Qualifiers
FT      1..1086
FT      CDS
FT      /*tag= a
XX
XX      MO9920766-A2.
XX
XX      29-APR-1999.
XX
XX      20-OCT-1998; 98WO-US22225.
XX
XX      15-MAY-1998; 98US-0085649.
XX      20-OCT-1997; 97US-0062592.
XX
XX      (GENZ ) GENZYME TRANSGENICS CORP.
XX
XX      Chen LH, Meade H;
XX
XX      WPI: 1999-302742/25.
XX      P-PSDB; AAY05833.
XX
XX      New modified recombinant nucleic acid sequences useful for producing
XX      malarial DNA vaccine
XX
XX      Disclosure: Fig 2: 43bp; English.
XX
XX      This nucleic acid encodes a 42 kDa C-terminal portion (see AAY05833)
XX      of malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
XX      important target for the development of a vaccine against
XX      Plasmodium falciparum. The 3' end of the sequence has been
XX      modified to include a 6xHis tag. The nucleic acid sequence has
XX      been modified (see AAX25586) according to a method of the invention
XX      in order to improve expression in mammalian host cells and in
XX      transgenic animals. In the modified coding sequence, 306 nucleotide
XX      positions have been replaced to lower the AT content (from 76 to
XX      49.7%) and to eliminate 10 mRNA instability motifs. The encoded
XX      amino acid sequence is unaltered. In another modified sequence
XX      (see AAX25593), a signal peptide sequence has been added and two
XX      N-glycosylation sites eliminated. The invention provides modified
XX      recombinant nucleic acid sequences and methods for increasing the
XX      mRNA levels and protein expression of proteins that are difficult
XX      to express in cell culture systems, mammalian cell culture systems
XX      or in transgenic animals. The preferred difficult protein
XX      candidates for expression are those derived from lower organisms
XX      such as parasites, bacteria and viruses that have DNA coding
XX      sequences of high AT content or which have mRNA instability motifs
XX      or rare codons relative to the recombinant expression system to be
XX      used. The invention allows expression of MSP-1 protein in the milk
XX      of transgenic animals, and also provides a DNA vaccine comprising a
XX      vector containing the altered MSP-1-42 sequence.
XX
XX      Sequence 1088 BP; 454 A; 139 C; 150 G; 345 T; 0 other;
Oy      3790 gcaagtaacccttcgtaattgataacacattctcaaatggaatgaaatgaatcaggtt 3849

```

```

|||||
Db 1 gcagtaactcctcgtaeltgtaacatacttctaaatgtgaaatgatacgggtt 60
3850 ttatattaaacctttagcaggtgttttagaagttaaaaaaacatagaataaac 3909
Db 61 ttatattaaacctttagcaggtgttttagaagttaaaaaaacatagaataaac 120
3910 gttatgacattttagtlttaagtgtagatattttaaattcacgatttaaacgtgaa 3969
Db 121 gttatgacattttagtlttaagtgtagatattttaaattcacgatttaaacgtgaa 180
3970 aatttcaaaaatgttttagaatacagatttaattccatataaagaatttaacaagtaac 4029
Db 181 aatttcaaaaatgttttagaatacagatttattccatataaagaatttaacaagtaac 240
4030 tatgtgtcaaaagatccatataaattctttaaagaagaagaagaataattccttaagc 4089
Db 241 tatgtgtcaaaagatccatataaattctttaaagaagaagaagaataattccttaagc 300
4090 agttataattatataaggaattcaatagatacagataaattctgcaaatgttctt 4149
Db 301 agttataattatataaggaattcaatagatacagataaattctgcaaatgttctt 360
4150 ggaattataaaaaatattatccgaaaaatataaatcagatttagatccaattaaataat 4209
Db 361 ggaattataaaaaatattatccgaaaaatataaatcagatttagatccaattaaataat 420
4210 atcaaacgacaaaaggtgtaaaatgagaatacttccttccttttaacaaatttgagac 4269
Db 421 atcaaacgacaaaaggtgtaaaatgagaatacttccttccttttaacaaatttgagac 480
4270 ttataaaaaacagtttaagtataaaattgtattttgttaattcatttagaagaagaagt 4329
Db 481 ttataaaaaacagtttaagtataaaattgtattttgttaattcatttagaagaagaagt 540
4330 ctaaatatatacatatgagaataatcaaacgttagaagaattaaagaacttaattctta 4389
Db 541 ctaaatatatacatatgagaataatcaaacgttagaagaattaaagaacttaattctta 600
4390 aaaaacattcaagacaatattgacagattttaaanaaaatacaaatctgttgatagct 4449
Db 601 aaaaacattcaagacaatattgacagattttaaanaaaatacaaatctgttgatagct 660
4450 gatttaacaagatataacataaacttaattgacaagaagtccttagaagaagt 4509
Db 661 gatttaacaagatataacataaacttaattgacaagaagtccttagaagaagt 720
4510 gttttgaaaattctgtctaaacggttttaattcttaatttactgtatgagaacttgcaagt 4569
Db 721 gttttgaaaattctgtctaaacggttttaattcttaatttactgtatgagaacttgcaagt 780
4570 atgttaaacatttcaacaacacacacacacacacacacacacacacacacacacacacac 4629
Db 781 atgttaaacatttcaacaacacacacacacacacacacacacacacacacacacacacac 840
4630 ttcaagacatttagatgaagaagaagaatgttaaatgttttaattatacaacaagaagt 4689
Db 841 ttcaagacatttagatgaagaagaagaatgttaaatgttttaattatacaacaagaagt 900
4690 gataaatgtgtggaataatccaaatcctactgttaacgaaataatgtgtgagtgtgca 4749
Db 901 gataaatgtgtggaataatccaaatcctactgttaacgaaataatgtgtgagtgtgca 960
4750 gttgtccaatgtgacgaaagaagaatctaggttagcagaagaagaagaataatcagtgatgt 4809
Db 961 gttgtccaatgtgacgaaagaagaatctaggttagcagaagaagaagaataatcagtgatgt 1020
4810 actaaacccgattcttatccactttcagatgtgattttcgcaggt 4854
Db 1021 actaaacccgattcttatccactttcagatgtgatttttctgcaggt 1065

```

RESULT 10

```

AAC68976
ID AAC68976 standard; DNA; 1077 BP.
XX
AC AAC68976;
XX
DT 27-FEB-2001 (first entry)
XX
DE Merozoite surface protein-142 coding sequence.
XX
KM Merozoite surface protein; protozoacide; vaccine; malaria; ss.
XX
OS Plasmodium falciparum.
XX
PN WO200063245-A2.
XX
PD 26-OCT-2000.
XX
PF 20-APR-2000; 2000MO-GB01558.
XX
PR 20-APR-1999; 99GB-0009072.
PR 13-MAY-1999; 99US-0311817.
PR 25-MAY-1999; 99CA-2271451.
XX
PA (MED-) MEDICAL RES COUNCIL.
XX
PI Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthairibull C;
DR WPI; 2001-015762/02.
XX
PT Novel variants of the C-terminal fragment of Plasmodium merozoite
PT surface protein-1, useful as vaccines for treating or preventing
PT malaria -
XX
PS Example 5; Fig 15; 126pp; English.
XX
CC The present invention relates to non-natural variants of a C-terminal
CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC non-natural variants have reduced affinity for at least 1 antibody
CC capable of blocking a second antibody that inhibits the proteolytic
CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC MSP-1.4.2, compared to natural MSP-1.1-9. The non-natural variants of the
CC present invention are useful for immunising a mammal against malaria, and
CC can be used to treat malaria. The present sequence is the MSP-142
CC coding sequence.
XX
SQ Sequence 1077 BP; 338 A; 228 C; 225 G; 286 T; 0 other:

Query Match 13.0%; Score 641.4; DB 22; Length 1077;
Best Local Similarity 74.8%; Pred. No. 1.4e-65;
Matches 804; Conservative 0; Mismatches 271; Indels 0; Gaps 0;

QY 3790 gcagtaactcctcgtaeltgataacatacttctaaatgtgaaatgataatgaggtt 3849
Db 1 gctgttaccatctgttatacgaatcttctgaatctgtctaaatgagaagaacagggc 60
QY 3850 ttatattaaacctttagcaggtgttttagaagttaaaaaaacatagaataaac 3909
Db 61 ttgtacttgaagccttgcggtgtgtctacagatcccttgaagaagaacactggaacacac 120
QY 3910 gttatgacattttagtlttaagtgtagatattttaaattcacgatttaaacgtgaa 3969
Db 121 gttatgacattttagtlttaagtgtagatattttaaattcacgatttaaacgtgaa 180
QY 3970 aatttcaaaaatgttttagaatacagatttaattccatataaagaatttaacaagtaac 4029
Db 181 aatttcaaaaatgttttagaatacagatttattccatataaagaatttaacaagtaac 240
QY 4030 tatgtgtcaaaagatccatataaattctttaaagaagaagaagaataatccttaagc 4089
Db 241 tatgtgtcaaaagatccatataaattctttaaagaagaagaagaataatccttaagc 300

```

[illegible]

PF 20-OCT-1998; 98WO-US222226.
PR
XX 15-MAY-1998; 98US-00856649.
PR 20-OCT-1997; 97US-0062592.
XX
XX PA (GENZ) GENZYME TRANSGENICS CORP.
XX
PI Chen LH, Meade H;
XX
XX WPI: 1999-288313/24.
DR P-PSDB; AAY09372.
XX
XX Modified malarial protein for use in anti-malarial vaccines
PT
PS Claim 1; Fig 1; 35pp; English.

This novel, modified nucleic acid encodes the 42 kDa C-terminal portion (see AA09372) of malaria merozoite surface protein MSP-1 (MSP-1-42), an important target for the development of a vaccine against Plasmodium falciparum. The nucleic acid sequence has been modified compared to the native sequence of MSP-1-42 (see AAX56009) such that 306 nucleotide positions have been replaced to lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA instability motifs while maintaining the same protein amino acid sequence. These alterations allow MSP-1-42 to be expressed in mammalian cell culture and in transgenic mice. Native MSP-1-12 is known to be difficult to express in cell culture systems, mammalian cell culture systems or in transgenic animals. The invention allows expression of MSP-1-42 protein in the milk of transgenic animals, and also provides a DNA vaccine comprising a vector containing the altered MSP-1-42 sequence.

Sequence 1065 BP; 324 A; 254 C; 280 G; 207 T; 0 other;

OY	4630	ttcgacatttagatgaagaagagttaatttataatttaacaagaaggt	4689
Db	841	ttcgcagatctcgacggagagaagaatgtaagtgtcttgaactaaacaggaaagt	900
OY	4690	gataaatgtttgaaaatccaatctctacttlaacgaaataatgtgtgatgca	4749
Db	901	gataagtggttgagaaccaccaacctactctttaacgagaacaacggtgatggaagct	960
OY	4750	gatgccaatgtacccgaagaagatctaggatgacaacggagaanaatcacatgtgaatgt	4809
Db	961	gacctaagtgccacggagaagaagacttgttcttaacggagaagaagattacttgcgaatgt	1020
OY	4810	actaaccttatcttatccacttttcgattgttatcttcgagttcctctaact	4864
Db	1021	actaagccagactcttacaccttgttcgatatgaactctctgtcttccctctaact	1075
 RESULT 11 AAx56008 ID AAX56008 standard; cDNA; 1065 BP. XX XX AAX56008; AC XX DT 31-AUG-1999 (first entry) XX DE Merozoite surface protein MSP-I-42 modified cDNA. XX KM MSP-I; merozoite surface protein; malaria; vaccine; KW protein engineering; codon usage; transgenic animal; ss. XX OS Plasmodium falciparum. XX Synthetic. OS PN WO9920774 -A2. PD 29-APR-1999. XX			

[illegible]

Db 481 ctgtacaagacgcgtcaacgataagattgattctgttcgtgataccacctggaagccaaggtc 540
Qy 4330 cttaattatacatatgagaataccaacgttagaatttaataaagaacttaattactta 4389
Db 541 ctgactacacacatctgagaagagcaacgttgaggttcaagatcaagagcctgtaattactt 600
Qy 4390 aaaaacttaacaagaacttggcagatttaaaaaataacaacttcgttggaaattgct 4449
Db 601 aagacccctcaagataagctgagcagatttcaagaagaacacacacttcgtcggatcgc 660
Qy 4450 gattatcaacagatttaacccaataaacttattgcaaaagtcccttagtaacagatg 4509
Db 661 gactcggacacgatttaacaacacaacacctgctgcaagttcctggaacacccgtatg 720
Qy 4510 gtcttgaataacttgcataaacaccttcttaacttaattacttgaatgaacttgaagt 4569
Db 721 gtctcgaataaacctgcaacagcgtctctgagcaacctgtctgtagtggaacctgcaagg 780
Qy 4570 atgttaacacttcaacaacccaatgcgtataaaaaaacaatgccaacaaattctgagt 4629
Db 781 atgctgacacatcagccagccagctgtgtgaagaagcagtgctccccaagaacagcgtgt 840
Qy 4630 ttcaagacatttagatgagaagaagaatgttaaatgtttatttaattacaacaagaagt 4689
Db 841 ttcaagacacctgtagagagagagagtgcaagtgctgtcgtgaacttaacagcaggaagt 900
Qy 4690 gataaattgtgtgaataccaatccctactgttaacgaataaattgttgatgtgca 4749
Db 901 gataaagtgtgtggaataaaccccaatccctactgttaacgaataaattgttgatgtgcc 960
Qy 4750 gatgccaatgtaccgaagaagattcagtagcaacggagaagaataacatgtgaatt 4809
Db 961 gatgccaagtgtaccgagagagattcagggagcaacggagaagaatcacacctgtgaatt 1020
Qy 4810 actaacacctatcttataccactttcagatgtgattttccgagct 4854
Db 1021 accaagcctgattctataccacgttcgatacgttcttctgtagt 1065

RESULT 12
AAK25586
ID AAK25586 standard; cDNA: 1065 BP.
AC AAK25586;
XX
XX
XX 02-AUG-1999 (first entry)
DE Merozoite surface protein MSP-1-42 modified cDNA.
XX
XX
XX MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal; ss.
XX
XX
XX Plasmodium falciparum.
OS Synthetic.
XX
XX
XX WO9920766-A2.
XX
XX
XX 29-APR-1999.
XX
XX
XX 20-OCT-1998; 98WO-US22225.
XX
XX
XX 15-MAY-1998; 98US-0085649.
XX
XX
XX 20-OCT-1997; 97US-0062592.
XX
XX
XX (GENZ) GENZYME TRANSGENICS CORP.
XX
XX
XX Chen LH, Meade H;
XX
XX
XX WPI: 1999-302742/25.
XX
XX
XX P-PsDB: AAY05832.
XX
XX
XX New modified recombinant nucleic acid sequences useful for producing

PT malarial DNA vaccine
XX
XX
XX Claim 12; Fig 1; 43pp: English.
CC This novel, modified nucleic acid encodes a 42 kDa C-terminal
CC part (see AAY05832) of malaria merozoite surface protein MSP-1
CC (MSP-1-42), an important target for the development of a vaccine
CC against Plasmodium falciparum. The nucleic acid sequence has been
CC modified compared to the native sequence of MSP-1-42 (see AAY25587)
CC such that 306 nucleotide positions have been replaced to lower the
CC AT content (from 76 to 49.78%) and to eliminate 10 mRNA instability
CC motifs while maintaining the same protein amino acid sequence.
CC These alterations allow MSP-1-42 to be expressed in mammalian cell
CC culture and in transgenic mice. The invention provides modified
CC recombinant nucleic acid sequences and methods for increasing the
CC mRNA levels and protein expression of proteins that are difficult
CC to express in cell culture systems, mammalian cell culture systems
CC or in transgenic animals. The preferred difficult protein
CC candidates for expression are those derived from lower organisms
CC such as parasites, bacteria and viruses that have DNA coding
CC sequences of high AT content or which have mRNA instability motifs
CC or rare codons relative to the recombinant expression system to be
CC used. The invention allows expression of MSP-1 protein in the milk
CC of transgenic animals, and also provides a DNA vaccine comprising a
CC vector containing the altered MSP-1-42 sequence.
XX
XX
XX Sequence 1065 BP; 324 A; 254 C; 280 G; 207 T; 0 other;

Query Match 12.2%; Score 601; DB 20; Length 1065;
Best Local Similarity 72.8%; Pred. No. 5,6e-61;
Matches 775; Conservative 0; Mismatches 290; Indels 0; Gaps 0;

Qy 3790 gcaatgactcctccggaattgataatcttctcaaatgtgaatgatggtt 3849
Db 1 gccgtacactccctccgacatgaataatcttccaagaatcgagaacgtacgagtg 60
Qy 3850 ttatattaaacctttagagagtggtttatagaagttttaaanaaacaattagaataac 3909
Db 61 ctgtactcgaagcgcgtcggaaggggtctaccgagacctgaagaacagcgtggaacaac 120
Qy 3910 gttatgacatttaagttaagttaagatatttaaatcagacttlaataacgtgaa 3969
Db 121 gtgtagaccttaacgtgaagcgtgaagatatacttgaaacgccgttcaacaagcggag 180
Qy 3970 aattcaaaaatttttagatcagatttaattccatataaagatttaacatgaatt 4029
Db 181 aacttaagaacgtgtcgagagcgtatctgataccctacaagaagatcgtacacgaac 240
Qy 4030 tatgtgtcaagaatcacaataaattcttaataagaanaaagagataaattcctaagc 4089
Db 241 taagtgtcgaagatccctcaagaagttccctgaacaagagagagataagttcccgagc 300
Qy 4090 agttataatataatgaagattcaatagatagacgataaatttgcacaatgttctt 4149
Db 301 agttacaactatacaagaatagatgactgtatccgataaacttcgcaagtgctcg 360
Qy 4150 ggaattataataatattatccgaaataataatcagatttagattcaatataaatt 4209
Db 361 ggaattataaagatccctgtcggagaagtaagaagcagatcgtatcaatcaagaagac 420
Qy 4210 atcaagacaanaaaggtgaanaatgagaatcctccctttaaanaaatttggagc 4269
Db 421 atcaagacaagaag 480
Qy 4270 ttatataaagcgttlaatgaataaattgtatttgaattcaatttaagaacaaagt 4329
Db 481 ctgtacaagacgctcaacgataagattgtgtcgtgataccacctgtgagccaaggtc 540
Qy 4330 cttaattatacatatgagaataccaacgttagaagtttaaaataaagaacttaattactta 4389
Db 541 ctgaactacacatctgagaagaacagcgtgaggttcaagaatcaagaagagtgattacctg 600

[illegible]

Dk

07

Dik

Oy

Dk

Oy

Dh

Ox

Dk

Q5

De

Q4

Dh

QY

Dib

५०

Dh

QY

Dh

Q5

Dh 703 gccgacatcgtgacccgalltaacacacaacacctgtgacccaagtctctgagcccgaa 762
Qy 4507 atggttttaaaatctgttaaaacccgtttatctaatcttgaatggaacttgcaa 4566
Dh 763 atggtcttgaaaaacccgtgacgaacccgtctctgagcaacctgtctgttgaaacttcag 822
Qy 4567 ggtatgttaaacatltcacacaacccaatgctglaaaaaaacaatgcccacaaatcttga 4626
Dh 823 ggaatgtgacagatcacgcacgacccaagtgtctgaaagaagcagtgctccacgaacgcyga 882
Qy 4627 tgtttgacacatttaagtgtgaagaagaatgtcaaatgtttatataattcaacaagaa 4686
Dh 883 tgccttcgaacacccgtgagtgaggagtgcaagtgcctgtcgaactacaagcagga 942
Qy 4687 ggtgataaaagtgttgaanaatcccaatccactgttaacgaaaataatgtgtgagtgat 4746
Dh 943 gggagataagtggttggaaaaaccccaatccactgttaacgaaaataatgtgtgagtgat 1002
Qy 4747 gccagatgcacaatgttaccgagaagaatcagtagcaacggaagaagaatacatgtgaa 4806
Dh 1003 gccagatgcacaatgttaccgagaagaatcagtagcaacggaagaagaatacatgtgaa 1062
Qy 4807 tgtacttaaacctgtattctatccacttttcgattgtgttttttcgcact 4854
Dh 1063 tgtaccaagcctgtattctatccactgttctgattgtatttttcgcact 1110

RESULT 14

AAx25593
ID AAx25593 standard; cDNA; 1140 BP.

AC AAx25593;

DT 02-AUG-1999 (first entry)

DE Merozoite surface protein MSP-1-42 modified cDNA.

KM MSP-1: merozoite surface protein; malaria; vaccine;

KW protein engineering; protein expression; codon usage;

KX transgenic animal; mutant; ss.

OS Plasmodium falciparum.

OS Synthetic.

XX Key Location/Qualifiers

FT CDS 1..1131

FT sig_peptide /tag- a

FT mat_peptide 46..1128

FT /tag- c

PN WO920766-A2.

PD 29-APR-1999.

PF 20-OCT-1998; 98WO-US22225.

PR 15-MAY-1998; 98US-0085649.

PR 20-OCT-1997; 97US-0062592.

XX (GEN2) GENZYME TRANSGENICS CORP.

XX Chen LH, Meade H;

XX WPI; 1999-302742/25.

XX P-PSDB; AAY05834.

XX New modified recombinant nucleic acid sequences useful for producing

XX Claim 12; Fig 11; 43pp; English.

CC This novel, modified nucleic acid encodes a 42 kDa C-terminal
CC part (see AAY05834) of malaria merozoite surface protein MSP-1
CC (MSP-1-42), an important target for the development of a vaccine
CC against Plasmodium falciparum. The nucleic acid sequence has been
CC modified compared to the native sequence of MSP-1-42 (see AAx25587)
CC such that the AT content has been reduced and 10 mRNA instability
CC motifs eliminated while maintaining the protein amino acid sequence.
CC In addition, a sequence encoding a 15-amino acid beta-casein signal
CC peptide has been added to the 5' end of the sequence, and N262Q
CC and N181Q mutations have been introduced to eliminate
CC N-glycosylation sites. These alterations allow MSP-1-42 to be
CC expressed in the mammary gland (i.e. milk) of transgenic mice. The
CC invention provides modified recombinant nucleic acid sequences and
CC methods for increasing the mRNA levels and protein expression of
CC proteins that are difficult to express in cell culture systems,
CC mammalian cell culture systems or in transgenic animals. The
CC preferred difficult protein candidates for expression are those
CC derived from lower organisms such as parasites, bacteria and
CC viruses that have DNA coding sequences of high AT content or which
CC have mRNA instability motifs or rare codons relative to the
CC recombinant expression system to be used. The invention allows
CC expression of MSP-1 in the milk of transgenic animals, and also
CC provides a DNA vaccine comprising a vector containing the altered
CC MSP-1-42 sequence.

Sequence 1140 BP; 353 A; 282 C; 290 G; 215 T; 0 other;

Query Match 12.0%; Score 588; DB 20; Length 1140;
Best Local Similarity 71.9%; Pred. No. 1,7e-59;
Matches 768; Conservative 0; Mismatches 300; Indels 0; Gaps 0;

Qy 3787 gaagcagtaactctctccgaattgataacatactcttcaaaattgaattgaattgag 3846
Dh 43 gccgcgcgaactccctccgcgaatcgaataacatccctgcgaagatgagagagtcag 102
Qy 3847 gtttataattaaacatttagcaggtgtttatagaagtttaaaaaacacattagaana 3906
Dh 103 gtctgtgaacttgagcccccgcagagagtcctacagagccttgaaagagcgcggagac 162
Qy 3907 aacgttaacatttaagttaagttaagataatttaattcaacgatttaataaagt 3966
Dh 163 aacgttaacatttaagttaagttaagataatttaattcaacgatttaataaagt 222
Qy 3967 gaaattcaaaaagtgtttagaatcagattatccatataagaatttaacataagt 4026
Dh 223 gagaacttcaagaacgtgctggagagcgatctgacccctcaagaagatctgacacagc 282
Qy 4027 aattatgtgtcaaaagatccataaaattcttcaataaagaagaagaataatctta 4086
Dh 283 aactagtggtcaaaagatccctcaaaattcttcaataaagaagaagaagaataatctta 342
Qy 4087 agcagttataattatataagaattcaatagatagatagataaaatttgcaaatgagt 4146
Dh 343 agcagttataattatataagaattcaatagatagatagatagatagatagatagat 402
Qy 4147 ctggatattataaataattatccgaaaataataaataagaatttaataaataa 4206
Dh 403 ctggatattataaataattatccgaaaataataaataagaatttaataaataa 462
Qy 4207 tataatcaagacaagaagtgaaaataagaataatccctctttaaacaatatgag 4266
Dh 463 tatatcaagataagcag 522
Qy 4267 acccttataaacaagtgaaatgaattgtatctgttaattcaattagaagaaca 4326
Dh 523 acccttataaacaagtgaaatgaattgtatctgttaattcaattagaagaaga 582
Qy 4327 gtcttaaatatcatatgagaatacaacgtagaggttaaaataaagaacttaattc 4386
Dh 583 gtcttgagtaacacataltgagaagaacagctgagagtgtaacagatcaagagctgaattc 642
Qy 4387 ttaaaacaattcaagaacaattgagatlttaaaaaataacaattcgttgat 4446

```

Db      643  ctgaagaccatccaggatgaagctgcccatttcagaagaacacacttcgcggaac 702
Qy      4447  gctgattatcaacaatataacataaacttattgaacaagtctctgaacagt 4506
Db      703  gccgactctgagccacattacaacacacactctgccaagtctcgaacccgga 762
Qy      4507  atgcttttgaataactcttcaaaccttttacttaattacttgatggaacttga 4566
Db      763  atgcttctgaaaactctgccaagacgcttcctgaacactctgtgatggaactc 822
Qy      4567  ggtatgttaaacatttcacaacacccatgctgaataaaacatgctcacaaatc 4626
Db      823  ggaatctctgagacgacgacgacgagctgtgtgaagaagagctcccccgaac 882
Qy      4627  tgttttaagcatttaagataagaagaagtgaatgttatttaattacaacaaga 4686
Db      883  tgcctcagacacactgtagagagaggtgcaagctgctggaactacaagcagaa 942
Qy      4687  ggtgaataatgtgtgaataatccaactcctactgtgaacgaataatgtgtgat 4746
Db      943  ggaagataagtgtgtgtaaaaaaccccaactctactgttaacgagaacatgga 1002
Qy      4747  gcagatgccaatgtatccgaagaagatcaggtagcaacggaagaacatcatgt 4806
Db      1003  gccagatgccaagtgtacgagagagatcagagaacggaagaagatcacctgca 1062
Qy      4807  tgtactaaacctgattctatccacttccgattcgtatcttcgcagt 4854
Db      1063  tgtaccagacctgattctatccacttccgattcgtatcttcgcagt 1110

```

RESULT 15

```

AA81151
ID      AA81151 standard; DNA; 1068 BP.
XX
AC      AA81151;
XX
DT      11-NOV-1990 (first entry)
XX
DE      DNA sequence encoding polypeptide p190-3.
XX
KW      Polypeptide p190-3; P.falciparum; merozoite; vaccine.
XX
OS      Plasmodium falciparum.
XX
FH      Key      Location/Qualifiers
FT      CDS      1..1068.
FT      CDS      /*tag= a
XX
PN      EP283829-A.
XX
PD      28-SEP-1988.
XX
PF      08-MAR-1988; 88EP-0103564.
XX
PR      19-MAR-1987; 87GB-0006599.
XX
PA      (HOFF) F. HOFFMANN-LA ROCHE & CO.
XX
PI      Ulrich C, Gentz H, Takacs B;
XX
DR      MPI: 1988-272339/39.
DR      P-PSDB; AAP80547.
XX
PT      New immunogenic polypeptides - derived from Plasmodium falciparum 190 kD
PT      surface antigen precursor, useful in vaccines, and encoding DNA sequences
XX
PS      Disclosure; 5 spp. German.
XX
CC      The DNA encodes a region of a polypeptide. The region contains at least
CC      one epitope of the 190kD precursor of the major merozoite surface antigen
CC      of P.falciparum.

```

```

XX
SQ      Sequence 1068 BP; 471 A; 129 C; 161 G; 307 T; 0 other;
Query Match      10.9%; Score 535; DB 9; Length 1068;
Best Local Similarity 87.9%; Pred. No. 1.8e-53;
Matches 583; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
Qy      331  acaatccctcagatattcaagtgatcagatcgaactcaacttcaagctgattacaac 390
Db      298  ataaaaataaaattatataaggtlaattcccaagcgtaataatacggattgaacat 357
Qy      391  agatcgaataactctgttaactcaagaagaactcaaatatcccaactcttgatta 450
Db      358  gcatatgaatcttcaaaaaaattcccccagaagagacgattgtcacaagttgaa 417.
Qy      451  actaatcatgttcaacttctgtgtgaatataatcatggtttcaaatatttaattga 510
Db      418  gaaagtgcacccgaacttctgtgtgaatattcatcattcgtttcaaatatttaattga 477
Qy      511  tatgaagaattatgaattatataataataaactttattttattataagaga 570
Db      478  tatgaagaattatgaattatataataataaactttattttattataagaga 537
Qy      571  aaatgaatgatgtatgtcttaattatgttcaaatcccttcaacttcaaatctgt 630
Db      538  aaatgaatgaatgtatgtcttaattatgttcaaatcccttcaacttcaaatctgt 597
Qy      631  gcaatgaattagacgtacttaaaaaactgtgttcggatatagaagaacatagaaca 690
Db      598  gcaatgaattagacgtacttaaaaaactgtgttcggatatagaagaacatagaaca 657
Qy      691  ataaagaataatgaagaagaagaagattacatttaaaaaataaaaaaccataga 750
Db      658  ataaagaataatgaagaagaagaagattacatttaaaaaataaaaaaccataga 717
Qy      751  aatataatgaatttaattgagaagaagtaagaacaacttgataaaaaataagaatga 810
Db      718  aatataatgaatttaattgagaagaagtaagaacaacttgataaaaaataagaatga 777
Qy      811  aaagaagaagaagaagaagaagaatataccaagctcaatatgacttctcatattaca 870
Db      778  aaagaagaagaagaagaagaagaatataccaagctcaatatgacttctcatattaca 837
Qy      871  caattgaagaagacacataatttaataagcttttgaaaaaagcttgaacactttaa 930
Db      838  caattgaagaagacacataatttaataagcttttgaaaaaagcttgaacactttaa 897
Qy      931  aaaaatgaagaacattgaagaacttgaataagaataatgaataaaatccccacccg 990
Db      898  aaaaatgaagaacattgaagaacttgaataagaataatgaataaaatccccacccg 957
Qy      991  gcc 993
Db      958  gcc 960

```

Search completed: October 27, 2001, 15:51:48
 Job time: 18444 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2001, 09:19:24 ; Search time 8587.58 Seconds
(without alignments)
8861.789 Million cell updates/sec

Title: us-09-269-874-1
Perfect score: 4920
Sequence: 1 atgaagatcatcttctt.....tatatacagttcatttaa 4920

Scoring table: IDENTITY-NUC
Gapop: 10.0, Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_in4:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_higo_hum:*
20: em_higo_inv:*
21: em_higo_rod:*
22: em_higo_hum1:*
23: em_higo_hum2:*
24: em_higo_hum3:*
25: em_higo_hum4:*
26: em_higo_hum5:*
27: em_higo_hum6:*
28: em_higo_hum7:*
29: em_higo_hum8:*
30: em_higo_inv1:*
31: em_higo_inv2:*
32: em_higo_inv3:*
33: em_higo_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_hum8:*
42: em_in:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_rod:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v1:*
59: gb_v12:*
60: gb_hgt1:*
61: gb_hgt2:*
62: gb_hgt3:*
63: gb_hgt4:*
64: gb_hgt5:*
65: gb_hgt6:*
66: gb_hgt7:*
67: gb_hgt8:*
68: gb_hgt9:*
69: gb_hgt10:*
70: gb_hgt11:*
71: gb_hgt12:*
72: gb_hgt13:*
73: gb_hgt14:*
74: gb_hgt15:*
75: gb_hgt16:*
76: gb_hgt17:*
77: gb_hgt18:*
78: gb_hgt19:*
79: gb_hgt20:*
80: gb_hgt21:*
81: gb_hgt22:*
82: gb_hgt23:*
83: gb_hgt24:*
84: gb_hgt25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_rol:*
95: gb_rol2:*
96: gb_rol3:*
97: gb_rol4:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	4920	100.0	4920	9	A92450	A92450 Sequence 1
2	4920	100.0	5917	96	PFANR195	X02919 Plasmodium
3	4920	100.0	7038	96	PFGR195A	X15063 Plasmodium
4	4779.2	97.1	5760	9	A04562	A04562 P.falciparum
5	4779.2	97.1	5760	10	E00656	E00656 cDNA encodi
6	4750.4	96.6	5282	96	PFPI90	X03371 P.falciparu
7	2853.4	58.0	5271	5	AF062349	AF062349 Plasmodu
8	2814.4	57.2	4940	9	A92451	A92451 Sequence 2

PD = 10/2/97

[illegible]

QY	61	acacatgaagcttatccagaagactgtccaaaaaacctagaagctttagaagaatgcagattg	120
Db	61	ACACATGAAGCTTATCCAGAACTGTCCAAAAAACAAGAGCTTTAGAAGATGCAAGTATTG	120
QY	121	acaggttatagtttatttccaaaaagaaaaatggtatttaaatgaagagaagaatgagaca	180
Db	121	ACAGGTTATGTTTATTTCCAAAAGAAAAAATGTTAATTAATGAAGGACAAAGTGGACA	180
QY	181	gcgtttacaactagtaacacctggttccaagaagttcaagttgcctcaagtggttcaagtggc	240
Db	181	GCCTTTACAACTAGTACACCTGGTTCCAAAGGTTGACGTTCAGGTTCCAGGTGCGC	240
QY	241	tcaagttgcttcagtggtgcctcagttgcttcacgttggtcctcaagttgcttcagttcagtt	300
Db	241	TCAGTTGCTTCAGGTGCGCTCAGTTGCTTCAGGTGCGCTCAGTTGCTTCAGGTGCGCTCAGTT	300
QY	301	gcttcaggttggtttcaggttaattccaagaagtcacaatcccttcagalaattccaagttgatca	360
Db	301	GCTTCAGGTGgtttcAGGTAAATTCAGAAAGCAATCCCTCAGTAATTCACAGTGATTCGA	360
QY	361	gatgcataaacttcaagctgatttaaaacacagagtagaatacttgcttgtaactatccaaa	420
Db	361	GATGCTAAATCTTACGCTGATTTAAACACAGAGTACGAATTACTTGTTAACTTATCCAAA	420
QY	421	gaactcaaatatccccaacctcttgatttaacctaatcatatgtaacttggtgataat	480
Db	421	GAACCTCAAAATATCCCAACCTCTTGATTTAATCATATATGTTAATGTTTGTGTGTAAT	480
QY	481	attcagtgcttcaaatatttaattgagatagagaagaattatgaattatataataa	540
Db	481	ATTCAATGCTTCAAAATTTAATTCATGAGATATGACAAATTAATGATTAATTAATTA	540
QY	541	ttaaaccttatttgatttatttaagaagcaaatlaaataatgataatgataatgataat	600
Db	541	TTTAAACTTTAATTTGATTTATTAAGAGCAAAATTAATATGATGATGCTCATATGATAT	600
QY	601	tgtcaaatacccttcaactcttaaaactcgtgcgaatgaattgaagttactttaaact	660
Db	601	TGTCAAAATACCTTTCAATCTTAAATTCGTGCATAATGATACGCTACTTAAAAACTT	660
QY	661	gtgttcggtatagaaaaaacatagacacatataaagataatgaaagaaaatggaagat	720
Db	661	GGTTCGGATATGAAAAACCATTTAGACAAATTTAAAGATTAATGTAGCAAAATGGAAGAT	720
QY	721	tacattaaaaaaaataaaaaaacatagaagaatataatgaatttaattgaagaagtaag	780
Db	721	TACATTTAAAAAATTAATAAAACCATTGAAAAATTAATTAATTAATTAATAAAGATAG	780
QY	781	aaacaacttgaataaaataagaaatgcacccaagaagaagaaaaaaaactataccaa	840
Db	781	AAAACAATTTGATTAATAAATTAAGAAATGCACTTAAGAAAGAAAGAAAAAATTAATACAA	840
QY	841	gtcaaatatgctcttcttatttacaataaaacaattagaagaagcacataatttaataagc	900
Db	841	GCTCAATATGCTCTTCTTATTTACATTAACAATTAAGAAAGCACTTAATTAATTAAGC	900
QY	901	gttttagaaaaacgttatctgaccttaaaaaaaatggaacacataagaagattactgat	960
Db	901	GTTTtagAAAAAGSTATTGCACTTTAAAAAATAATGAACAATTAAGGAATTAATCTTGAT	960
QY	961	aagataatgtaaatlaaaaaatccccccacccggccaattctggaataataccaaatacttc	1020
Db	961	AAGATAATGTAATTAATAAATCCCCCACCGGCCAATTTGCAAAATATACCAAAATTAATCTTC	1020
QY	1021	cttataaagaacaaaaaattcgaggaacgcgaaaaagaattataaagaattgccaact	1080
Db	1021	CTTATATAGAACAAAAAATCGAGGACACGAAAAAGAAATTAAGAAATTTGCCAAAATCT	1080
QY	1081	attaaatttaatatlgatagtttatttacttgaatgcaactggaatagaataactattaga	1140
Db	1081	ATTAAATTTAATATGATAGTTAATTAATTAATGATGCACCTTGGAATTAAGATTAATTAAGA	1140

OY	1141	gaaaaaataaaaaatcttgatctaagtcgaaggttcgaacaaagaaatcaactgaaacc	1200
OY	1141	gaaaaaaatpaaaaatpattgatpaaagtcgaaaggttgaaaacaaagaaatcaactgaaacc	1200
OY	1201	aatgaatattccaaatcgagatgaactatcccttgctatgaataaagatattaaagcttga	1260
Dp	1201	aaagaatattccaaatgaagattgaactatcccttgctatgaataaagatattaaagcttga	1260
OY	1261	aatgaacttaattcctcttggtgatctaaatcaatccaatcttgatctaaagaacgaagt	1320
Dp	1261	aaatgaacttaattcctcttggtgatctaaatcaatccaatcttgatctaaagaacgaagt	1320
OY	1321	aaaaacatatatctgataacgaaagaaaaatctcaataatgaatgaagaaaaat	1380
Dp	1321	aaaaacatatatctgataacgaaagaaaaatctcaataatgaatgaagaaaaat	1380
OY	1381	aaaaatgaaaaaaaaatctgaatctgaataaaaaatcttcgaagacgaatctgaatct	1440
Dp	1381	aaaaatgaaaaaaaaatctgaatctgaataaaaatcttcgaagacgaatctgaatct	1440
OY	1441	ctaaatgatatacaaaagaatatagaaaatctactaaatgaattatgatagaacatctc	1500
Dp	1441	ctaaatgatatacaaaagaatatagaaaatctactaaatgaattatgatagaacatctc	1500
OY	1501	aataataatagatcttaactcaatctcgaaaaatgatagtggtlaaagaatctcataaa	1560
Dp	1501	aaatpattatpattatgatttgaatttgaatttcgaaaaaaatgattggaagatttgaatt	1560
OY	1561	gttgagaactatcacacatatgaactcttgatctcccttgaaatcttaaacataactct	1620
Dp	1561	gttgagaactatcacacatatgaactcttgatctcccttgaaatcttaaacataactct	1620
OY	1621	gaaaagttaacaaagcctctaaataatgatgagaatactctcttaagaatatagtatgt	1680
Dp	1621	gaaaagttaacaaagcctctaaataatgatgagaatatttctttagaagaattatgattgt	1680
OY	1681	gaaaagaatctaaatatataataaaatcttaataagcaaaatagaatggaaatgaaaca	1740
Dp	1681	gaaaaaaatpaaaaatpattgatpaaaaatpattgaatpaaaaatgaaatgaaaca	1740
OY	1741	tttgttgaaaaatatctaaaaaagatgaaagaaagcctcttgaaaaaaaatcttaacgaac	1800
Dp	1741	tttgttgaaaaatatpaaaaaagatgaaagaaagcctcttgaaaaaaaatpactgaagac	1800
OY	1801	gaaaataaacccagatgaaaaaaatctttagaagatcttgacatgataaagtaaacgttcaa	1860
Dp	1801	gaaaatpaaacacacatgaaaaaaatctttagaagatcttgacatgataaagtaaacgttcaa	1860
OY	1861	aaagcttattataatgaaacaaatctgaacgaattaaagaagctcaatgattcttaaaat	1920
Dp	1861	aaagcttattataatgaaacaaatctgaaatgaaatgaaagacatgatttgaatttgaatt	1920
OY	1921	gttgaatattaaacatatatactgaatgtctcccaattcttcaacaaagaagaataaagaagaa	1980
Dp	1921	gttgaatattaaacatatatactgaatgtctcccaattcttcaacaaagaagaataaagaagaa	1980
OY	1981	cccttatattctaaatctgcttgaaaaaaagaatctgataaataaagttctcaatgcctaaag	2040
Dp	1981	cccttatattctaaatctgcttgaaaaaaagaatctgataaataaagttctcaatgcctaaag	2040
OY	2041	gttagaatctatctgataaattgaaagaaaaaaaacataaaaaacagaagttcaatcgataat	2100
Dp	2041	gttagaatctatctgataaattgaaagaaaaaaaacataaaaaacagaagttcaatcgataat	2100
OY	2101	tcggaaccatctcaaccgaagagagaataaaccagagcaagcaactataaaacctggaacacaa	2160
Dp	2101	tcggaaccatctcaaccgaagagagaataaaccagagcaagcaactataaaacctggaacacaa	2160
OY	2161	gcaagatctgctcttagaagagagatactgaatcaagaacaagacaagaacaaacaaagca	2220
Dp	2161	gcaagatctgctcttagaagagagatctgaatcaagaacaagacaagaacaaacaaagca	2220
OY	2221	caaacaccagattaccagatccagatcacagaagaacaaagcaagaatcccaacaccacagca	2280

Db	2221	CAACCCAGTACGAGTACCGATCCAGACGAAAGCAAGCAAGTCCCAACACCAAGCA	2280
Qy	2281	ccag taagataaagctggaagatgcttccaaatagatlatctggaagatlatgaa	2340
Db	2281	CCATTAATATATAAACTGAAATGTGTTCCAAATTAGATATCTTGAAAAATTATATGAA	2340
Qy	2341	tcttaaatcttcotabatalatgctcaaatatalttggttccactcaactlgaa	2400
Db	2341	TTTTTAATATCTTCATATATATGTGCAAAATATATTTGGTTTCCACCTCAACATGTAGAC	2400
Qy	2401	gaagaatattaacaactataaattcaagaagggaagaagaactiaagttcagtt	2460
Db	2401	GAAGAATATTAAACAAATATTAATAATTACAAAGAGAGAAAGCAAAATTAAGTTCACT	2460
Qy	2461	gacccatlagacttatgctttaaatacaaaataacatccgttaatglatctatgctt	2520
Db	2461	GATCCATTAGACTTATGTTTAATATATCAAAATATACATACGTATGTATGTATCTATGTT	2520
Qy	2521	gatagctttaaacaatagtttatacacaacttataygaaattatgaaaagaatggtt	2580
Db	2521	GATAGCTTTAAACATAGTTATTCACAACATATTTATGGAATTTATGAAAAAGAAATGGTT	2580
Qy	2581	tgtattatataaacttaaggataatgcaaatataaactttctttaggaagacgaa	2640
Db	2581	TGTAAATTTATATAACCTTAAGATATATGACAAATTAATAATTTATTTAGAGAACCGAA	2640
Qy	2641	aaagatccacactcgtgaanaaactcttcaagttcataatgcaaccatlatcaaca	2700
Db	2641	AAAGTATCCACACTGTGTAATAAACTCTTTCAAGTTTCATCAATGCAACACTATATCATTAACA	2700
Qy	2701	cctcaagataaaccggaagtaagtgcgaatgaigatatacscatcttcaaatltgaat	2760
Db	2701	CCTCAGGATTAACCCGGAAGTAAAGTGCAAATATGATGATCATCATCTTCACAAATTTGAT	2760
Qy	2761	aatgctttaaactatattgnaaacatatagagctctggaagaaacaaatatatacga	2820
Db	2761	AATGTTTTAAATTAATTTTGAAAACTATATGTAGCTTTGGAAGAAACCAAAATATATACCA	2820
Qy	2821	gaatataatgctcaaaaagtagtgaanaacttlatgaaagatlatlaaagatagttgat	2880
Db	2821	GAAATTAATAGCTCAAAAAGTAGGAANAACCTTTATGAAGAAACATATTAAGATAGTGAT	2880
Qy	2881	acatttataatgaatctttacaacttttgtaaaactcaagctgtagataataatca	2940
Db	2881	ACATTTTATATGAACTCTTTTACAAATTTGTAAATCTTAAGAGCTGTGATATTTATTTCA	2940
Qy	2941	ttgaatgtagatacaaaaaggagaatattgaagaagatattataaacttlaaaaaaact	3000
Db	2941	TTGAATGTATGAATCAAAAAGCAAAATTTGAAAGAAATTAATTAATAATTAATAAAAAACT	3000
Qy	3001	ttacagttatcatcttgattatataataaataaacttlaaagatlaagacttattgat	3060
Db	3001	TTACAGTTATCATTTGATTTTATATATAAATATATAATTTAAAAATTAACAAGATTATTTGAT	3060
Qy	3061	aaaaaagaacagtttggtaatatataaaygcaaatataaanaacttacttlatlaaaga	3120
Db	3061	AAAAAGAAACAGTTGGTAAATATAAATCAAAATTTAAAAAACTTCTTATTTAAAAAGAA	3120
Qy	3121	caattagaatacaaaattggaattcaacttaataaccacaagcatglatataaanaactttct	3180
Db	3121	CAATTAGAAATCAAAATTTGAATTCCTTAATAATACCCAAAGCATGTATTTACAAAATTTTCT	3180
Qy	3181	gttttccttacaagaaagagagctggaatagcagaanaactgaaaacatlagaaaac	3240
Db	3181	GTTTTCCTTTAAACAAAAAAGAAAGCTGGAATATGCAAGAACTGAAAAACATTTAGAAAAAC	3240
Qy	3241	acaaaaatataatlgaaacattataaagaactgttinaatbatataaygtygaatcatct	3300
Db	3241	ACAAAAATATTATTGAAACATTATAAGAGACTGTTAAATATATTAATGTGTAATCATCTT	3300
Qy	3301	ccattaaaaactttaagtgaagaatcaatccaagagaataataatgacagtttagaa	3360

Db 3301 CCATTAAAACTTTAAGTGAAGATCATTCATCAACAGAAAGATTAATTGCGCATTTAGAA 3360
 Oy 3361 aactttaaagtattaaattagaagaaattaaaggtaattaaattagaagaag 3420
 Db 3361 AACTTTAAAGTATTAAAGTAAATTGAAGAAATTTAAAGATTAATTAAATTGAGAAAG 3420
 Oy 3421 aaaaaatcatcactatcaagaatgagatcatcattcaattgctgtaataagaaga 3480
 Db 3421 AAAAAATTTCACTACTTATCAACAGTATGATACATCTTAATTTCTGCTGAATTAAGAAGTA 3480
 Oy 3481 ataaaaataaaaaataatcatcaggtaattctcccaagtgaaaaataatcagatgtaacaat 3540
 Db 3481 ATAAAAATAAAAAATTAATTAACAGTAAATTTCTCCAAAGTGAATAATATCGGTGTTAAACAAT 3540
 Oy 3541 gcaatagaactctacaaaaaatctccccaagaagacagatgctgcaacagttgtaagt 3600
 Db 3541 GCATTGAATCTTACAAAAATTTCTCCCAAGAAAGAACAGATTTTCCACAGCTTGTAAGT 3600
 Oy 3601 gaaagtgcacgcacacattagaacaaagtcacaaagaacacagcatcaactcagta 3660
 Db 3601 GAAAGTGATCCGACACATTAGAACAAAGTCAACCAAGAACACAGCATCACTCATGTA 3660
 Oy 3661 ggagagagatctcaacacaaataacacatacacaagaatgctgagatgagatgagta 3720
 Db 3661 GGAGCAGAGTCTACACAAATTAACACATCAACAAATGTCATGATGAAGTAGACGTA 3720
 Oy 3721 atcatagtaactatatttgagaaatccgaagaagatcatcagatgattagaagaagta 3780
 Db 3721 ATCATAGTACCTATATATTTGAGATCCGAAGAAGATTATGATGATTTAGACAGTAGTA 3780
 Oy 3781 acaagagagaagcagtaactccctccglaattgatacattcttcaaaatgaaaaatga 3840
 Db 3781 ACAGAGAGAGCGAGTAACCTCTCCGTAATTTGATTAACATTAATTTCTTAAATTTGAAATGAA 3840
 Oy 3841 tatggagtttatatttaaacccttaagaaggtgtttatagaagtttaaaaaacacata 3900
 Db 3841 TATGAGTATTATTTAATTAACCTTTAGCAGGTGTTAATGAAGTTTAAAAAAACATTA 3900
 Oy 3901 gaaaaataacgtatgacatttaattgtaattgaagatattttaaattcacagattta 3960
 Db 3901 GAAATTAACGTATGACATTTAATGTTAATGTTAATGATTTAATTTAATTTACAGATTAAAT 3960
 Oy 3961 aaagctgaaaaattcaaaaatgctttagaatcagatttaattccatataagaattca 4020
 Db 3961 AAAGCTGAAAAATTTCAAAATGCTTTAGAACATGATTAATTTCCATATTAATTAATTAACA 4020
 Oy 4021 tcaagtaattatgcttcaagaatcatataaattcttaataagaagaagaagaataa 4080
 Db 4021 TCAAGTAAATATGTTGTCNAAGATCCATTAATTTCTTAATTAAGAAAAAGAGATTA 4080
 Oy 4081 ttcttaagcagttataatataatlaagaatcagatagatagatagaataatttccaat 4140
 Db 4081 TTCTTAAGCAGTTAATTAATTAATTAAGATTTCAATAGATTAAGATTAATTTTCCAAAT 4140
 Oy 4141 gatgttcttgatattataaaatlatatccgaaaaaataatcagatttagattcaat 4200
 Db 4141 GATGTCTTGGATATTATAAAATTAATTCGAAAAATTAATTAATTAATTTAGATTAATTAAT 4200
 Oy 4201 aaaaaatcatcaacgaacaaagaatgaaaaatgagaataactcccttttaacaat 4260
 Db 4201 AAAAAATTTATCAACGACAAACAGTGAATAATGAATAATACCTTCCCTTTTAAACAAT 4260
 Oy 4261 attgagacctataaacaagttatgataaaattgatttatttgtaattcatctaga 4320
 Db 4261 ATTGAGACCTATATATAAAGATTAATTAATTAATTTGATTTTGTAAATTCATTTAGAA 4320
 Oy 4321 gcaaaagtctcaaatcatatagagaatcaaaacgtagaagttaaaataagaactt 4380
 Db 4321 GCAAAAGTCTTAAATTAATTAATTAAGAAATCAAAACGTAGAAGTTAAATTAAGAACTT 4380
 Oy 4381 aattactaaaaacaattcaagaacaaattgacagattttaaataaaatacaaatctgt 4440
 Db 4381 AATTACTTAAAAACAATTTCAAGACAAATTTGACAGATTTTAAAAAAATTAACATTTTCGT 4440

Oy 4441 ggaattgcattatcatcaagaattataacacataaactattgcaaaagtccctagt 4500
 Db 4441 GGAATTGCGATTATTAACAGATTTAATTAACATTAATTAATTTGACAAAGTCCCTTAG 4500
 Oy 4501 acagatagtttctgaaactctgctaaacgctttatcattacttgaatggaac 4560
 Db 4501 ACAGATAGTGTTTTGAATAATCTTGCTAAACCGTTTATCTAATTTACTTGATGAAAC 4560
 Oy 4561 tgcagaatgattttaaactttcaacaacacaaatgcgtaaaaaaacaatgctcacaat 4620
 Db 4561 TTCAGAGTATGTTTAAACATTTTCACAACCAATGCGTAAAAAAACAAATGTCACAAAT 4620
 Oy 4621 tctgagatttgcagacattagaatgaagaagaatgtaaaatgattatataatcaaa 4680
 Db 4621 TCTGATGTTTTCAGACATTAGATGAAGGAAGATGTAATGTTTAAATTTGACAA 4680
 Oy 4681 caagaagtgatataatgatttgaanaatcccaatccctactgtaagaagaataatggtga 4740
 Db 4681 CAAGAAGTGATTAATGTTGTAATAATCCAAATCTTACTTGTACCAAAATTAATGTTGA 4740
 Oy 4741 tctgagcagatgacaaatgtaaccgaagaagatcagtagaagaacgaaagaaatcaca 4800
 Db 4741 TGTGATGACAGATGCCAAATGTAACCGAAGAAGATTCAGTAGCAACGAAAGAAATCACA 4800
 Oy 4801 tctgagatgtaacaaactgattctatccacttttcgattgatttctgcagttccct 4860
 Db 4801 TGTGATGTAACCTGATTTTATCCACTTTTCGATGATGTTTTCGAGTTCCTCT 4860
 Oy 4861 aacttctagaatcatcttattatcaactatgtaataatgtaataatgtaataatgta 4920
 Db 4861 AACTTCTTGAATATCATCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4920

RESULT 2

PEANT195 5917 bp mRNA INV 30-MAR-1995
 LOCUS
 DEFINITION Plasmodium falciparum mRNA for major surface antigens precursor (P195).

ACCESSION X02819.1 GI:9864
 VERSION
 KEYWORDS antigen; direct repeat; signal peptide; surface antigen.
 SOURCE malaria parasite P. falciparum.
 ORGANISM Plasmodium falciparum

REFERENCE 1 (bases 1 to 5917)
 AUTHORS Holder, A.A., Lockyer, M.J., Odink, K.G., Sandhu, J.S., Riveros-Moreno, V., Nicholls, S.C., Hillman, Y., Davey, L.S., Tizard, M.L.V., Schwarz, R.T. and Freeman, R.R.
 Primary structure of the precursor to the three major surface antigens of Plasmodium falciparum merozoites.

TITLE
 JOURNAL Nature 317 (6034) : 270-273. (1985).
 MEDLINE 8601435
 REFERENCE 2 (bases 1 to 5917)
 AUTHORS Holder, A.A.

JOURNAL Direct Submission

REMARK Submitted (06-MAR-1991) Holder A.A., National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK
 COMMENT Update of published sequence
 DATA kindly reviewed (06-MAR-1986) by R.R. Freeman.
 FEATURES
 Location/Qualifiers

source

sig_peptide
 418..474
 /product="put. signal peptide (aa -19 to -1)"
 418..5337
 /note="p195 precursor"

CDS

/codon_start=1
 /protein_id="CAA26676.1"
 /db_xref="GI:9865"
 /db_xref="SWISS-PROT:P04933"
 /translation="MKRIIFLCSPLFIINTGCVTHESYQELVKKLEALDENVLTGTS"


```
|||||
Db 1618 AATGAATATCCAAATGAGAGTACTTATCCCTTTCATATACGATATTTAAACAATGCTTTA 1677
QY 1261 aatgaactaatctcttcttggttgatlaattacatccatcttgatctacaagaaccagc 1320
Db 1678 AATGAACCTTAATTCCTTGGTGAATTAATTAATTCATTTGATTAATTCACAAAGAACCAAGT 1737
QY 1321 aaaaacatatactgataatgaagaanaaaatccataatgaattgaaggaanaatc 1380
Db 1738 AAAACATATATCTGATATATGAAGAAAAATTCATTAATGAATTAAGAAAAATTT 1797
QY 1381 aaaaatgaanaaaaaaaatctgataaanaaatctgaagaacagatctaatgtc 1440
Db 1798 AAAATGAAAAAATAAATTTGAATCTGATTAATAAATCTTACGACACATCTAAAGTCT 1857
QY 1441 ttaaatgataaacaagaatgaanaaaatcttaataaanaattatagtaagaatc 1500
Db 1858 TTAAATGATTAACAANAAGATATGAAAAATTTCTTAATGAATTAATGATGACCAATTC 1917
QY 1501 aataataataagatcttaacttaattcgaanaaatgaatggtaaaagatactcataa 1560
Db 1918 AATAAATATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1977
QY 1561 gttagaactgaacacacaaatctcttgcaatcctaagaanaatcttaacaatct 1620
Db 1978 GTTGACAAACTTACACACCATTAATCTTTGATCCATGAAAAATCTTAACATTAATCTT 2037
QY 1621 gaaaagttaacaaaagctcttaataatgaagaatctctcttaaggaatagtagt 1680
Db 2038 GAAAAATTAACAAGCTTTAATATATGGAAGATTAATCTTTAAGGAATTAATGATGCTT 2097
QY 1681 gaaaagaattaaatatataaanaattataagcaaaaatagaanaatgaatctgaaca 1740
Db 2098 GAAAAAGATTAATAATATTAATAAATTTTAATAGCAAAATTAAGAAATGATTAAGAAACA 2157
QY 1741 ttagttgaaaatcttaaaaaagatgaagaacgctcttgaaaaaanaatcttaagaac 1800
Db 2158 TTATGTTGAANAATTTTAAAAAAGATGAAGAACACCTTTTCAAAAAAATAATTAATAAGAC 2217
QY 1801 gaaaataaaccaagatgaanaaatcttgagaatgaatctgaacatctgaaagtaagaatc 1860
Db 2218 GAAAAATTAACCGATTAATAAATTTTGAAGATATCTGACATTTGTAANAAGTACAAAGTTCAA 2277
QY 1861 aaagtcttaataatgaacaaaatctgaacaaatlaaaaaagatcgaatcttaaaaaat 1920
Db 2278 AAGGTTTATTAATGAACAAATTTGACGAATTAATAAAGATCTCAATTTGATTTAAAAAAT 2337
QY 1921 gtagaatataaataataataatctcccaatctcttaacaagaanaataagcaaga 1980
Db 2338 GTAGAAATTAATAACATATATATCATGTTCCCAATTTCTTACAAACAAATAAATAAGCAAGAA 2397
QY 1981 ccttatctatcttaattgtgtgaaaaaagaatgtgataaattgaagttgttcagcttaag 2040
Db 2398 CCTTATTAATTAATTTGTGTTGAAAAAAGAAATTTGATTAATTAAGAGTTCATGAGCTTAAG 2457
QY 2041 gtagaatctatgaataatgaagaanaaaaaacataaaaaacagaagttcaatgagtaat 2100
Db 2458 GTAGATTCATTTGATTAATGAAGAAAAAATAAATAAATAAAGAGTCAATGAGTAAT 2517
QY 2101 tggagaacctcaacggaaggaataataacaggaacgaactcaaaactctgacaacaa 2160
Db 2518 TCGGAACCATCAACCGAAGGAATAATACGACGAACCAACTCAAAACCTTGACAAACAA 2577
QY 2161 gcaagatctgcctttagaagaagatctgaatcaagaacaaacacagaagaacaaagaac 2220
Db 2578 GCAGATTCCTTACAGAGAGATTTCAATGACACACACAGCAAGAACCAAAACAAAGCA 2637
QY 2221 caaccacgaatccagataccagataccagaagaacaaagcaaaatctcccaacacacagca 2280
Db 2638 CAACCAACGATCCAGTACACAGTACAGAAAGCAAAAGCAAAAGTCCCAACACACAGCA 2697
QY 2281 ccagtaataataaactgaanaatgtctccaaatagatctatctgaaaaatataatga 2340
|||||

Db 2698 CCAATAAATATAAATACTGAAAAATGTTTCCAAATTAATGATTAATCTGAAAAATTAATGA 2757
QY 2341 ttttaataactcatatataatgcacaataatcttggtcttcacacacacactgaac 2400
Db 2758 TTTTAAATTAATCTTCAATATATATATGTCACAAATATATTTTGGTTTCACCTCAACTATGAC 2817
QY 2401 gaaaagatataaacaataataaataacagaagaagaagaacaaatlaagttcaatgt 2460
Db 2818 GAAAGATTAATAAACATTAATAAATTTACAAAGAGGAAGAAAGCAAAATTAAGTTCATGT 2877
QY 2461 gatccatgaactatctgttlaatalacaaataacacacacacacacacacacacacac 2520
Db 2878 GATCCATTAATCACTTAATGTTTAATTAACAAATAATCAATCACTTAATGATTAATGAT 2937
QY 2521 gatagcttaacaaatgattatcacacacacacacacacacacacacacacacacacacac 2580
Db 2938 GATAGCTTAATCAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 2997
QY 2581 tgaattataataacttaagaataatgaacaaaatlaaaatcttaagaagaagcaaa 2640
Db 2998 TGTAATTTAATAAATTAATTAAGATTAATGAACAAATTAATAAATTTAATTAAGAGAGGAAA 3057
QY 2641 aaagtatcaacatctgtaaaaactcttcaagttcaatcaatgaacacacacacacacacac 2700
Db 3058 AAAGTATCCACATCTGTAATAAATCTTTCAAGTTCATCAATGCAACCATTAATCAATTAAC 3117
QY 2701 cctcagatataacccgaagaatgaacaaatgaatgaatgaatgaatgaatgaatgaatgaat 2760
Db 3118 CTTGAGATTAACCCGAGTAATGCAATGATGATGATGATGATGATGATGATGATGATGAT 3177
QY 2761 aatagttlaaatatctgaacacacacacacacacacacacacacacacacacacacacac 2820
Db 3178 AATAGTTTAATAATTTATTTGAAAAATTAATTAAGATTAATTAATTAATTAATTAATTAAT 3237
QY 2821 gaattaaatgaacaaaagatgaacaaactcttaagaagaatcttaagaagaatgaatgaat 2880
Db 3238 GAATTAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3297
QY 2881 acattataatgaatctcttatacaaatcttgtaaaatcttaaaagctgaatgaatgaatgaat 2940
Db 3298 ACATTTTATTAATTAATCTTTTACAAATTTTGTAAATCTTAACCTGATGATTAATTTCA 3357
QY 2941 ttgaatgaatgaacaaaaggaagaatlagagaagaatataataaataaataaataaataa 3000
Db 3358 TTGAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3417
QY 3001 ttagatctatctatctgattataataataataaataaataaataaataaataaataaataa 3060
Db 3418 TTACAGTTTATCAATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3477
QY 3061 aaaaagaagaacgcttgtaataataaataaataaataaataaataaataaataaataaataa 3120
Db 3478 AAAAAAGAAACAGTGTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3537
QY 3121 caattgaatcaaaatctgaatcttaataaataaataaataaataaataaataaataaataa 3180
Db 3538 CAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3597
QY 3181 gtctctcttaacaaaaaagaagaactgaataatgaagaacgtaaaaaacacacacacacac 3240
Db 3598 GTTTTCTTTAAACAAAAAAGAAAGAGCTGAATTAATTAATTAATTAATTAATTAATTAATTAAT 3657
QY 3241 acaaaaataatctgaacacacacacacacacacacacacacacacacacacacacacacac 3300
Db 3658 ACAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3717
QY 3301 ccaatlaaaaacttaagttgaagaatctcaacacacacacacacacacacacacacacacac 3360
Db 3718 CATTTAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3777
QY 3361 aacttaagaatlaagtaataatgaagaagaataaagaatlaagaatlaagaatlaagaatga 3420
Db 3778 AACTTTAAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3837
```

QY	4501	acaggtatggtttcttgaacaaactctgtctaaaccggtttacttaattacttga	tggaac	4500		
Db	4918	ACAGGTATGGTTTTGGAAAACTTGGTAAAAACCGTTTATCTAAATTTCTGTATGGAAAC	4977			
QY	4561	cttgaagatgatgttaaacatttcacacccaaatgtgtataaaaaacaatgtccaaaat	4620			
Db	4978	TTGCAAGGTATGTTAAACATTTTACACACCCAAATGGCTAAAAAAAACAATGTCCACAAAT	5037			
QY	4621	tctgtagtcttcagacacatttga	tgaaagagaagaa	tgttaatgttctaatactacaa	4680	
Db	5038	TCTGGATGTTTCACACATTATAGATGAAAGAGAAAGTAATGTATTAAATTACAA	5097			
QY	4661	caagaagtgatataa	tggtgttgaacatccaatccactctt	gtacgaaaaata	tgvtga	4740
Db	5098	CAAAAGGTGATAAATGTTGTGAAAAATCCAAATCCATCTTGTAAAGAAATATAGTGGA	5157			
QY	4741	tgtatgcagaatgtccaaatgtaccggaagaagtctcaggtagacaacgaaagaatacga	4800			
Db	5158	TGTGATGCAGATGCCAANATTAACCGAAGAGATTCAAGTAGACAGGAGAAATTCACA	5217			
QY	4801	tgtaatgtactaaacctgtatctctacactt	lccga	tggtatcttctgcagctccct	4860	
Db	5218	TGTGAATGTAACAACTGATCTTATCCACTTTTGATGGATTTTTCGAGTCTCT	5277			
QY	4861	aactcttgaagataatacttacttaataactca	tgvtataataataacagtttcaat	4920		
Db	5278	AACTTCTTAGGAATATCATCTTATAAATCACTCATGTTAAATTTATACGTTTCATTAA	5337			

```

RESULT 3
LOCUS PEGP195A 7038 bp mRNA INV 03-APR-1995
DEFINITION Plasmodium falciparum mRNA for major merozoite surface antigen
                    gp195.
ACCESSION X15063
VERSION X15063.1 GI:9896
KEYWORDS antigen; glycoprotein; gp195 gene; surface antigen.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
REFERENCE 1 (bases 1 to 7038)
AUTHORS Myler, P.J.
TITLE Direct Submmission
JOURNAL Submitted (14-APR-1989) Myler P.J., Seattle Biomedical Research
INSTITUTE, 4 Nickerson Street, Seattle WA 98109-1651, U S A
REFERENCE 2. (bases 1 to 7038)
AUTHORS Myler, P.J.
TITLE Nucleotide and deduced amino acid sequence of the gp195 (MSA-1)
JOURNAL gene from Plasmodium falciparum Palo Alto PLF-3/B11
MEDLINE Nucleic Acids Res. 17 (13), 5401 (1989)
FEATURES
         source            89345116
         location/Qualifiers
             1..7038
             /organism="Plasmodium falciparum"
             /strain="Palo Alto PLF-3/B11"
             /db_xref="taxon:5833"
             /dev_stage="trophozoite/schizont."
             1592..1648
             /product="signal peptide (AA -19 to -1)"
             1592..6511
             /codon_start=-1
             /product="gp195 surface antigen preprotein"
             /protein_id="CAA33163.1"
             /db_xref="GI:9897"
             /translation="MKTIIFLCSPFLFIIMTQCTHESYOELVKLEALDAVLTVGS
             LFOEKRAVLNKGSTGTAVTTPSGSKGSAVSGSGCSVAGSGSAVSGSVAS
             GCGSNRRTSPDSSDSDAKSVADIKHRPARYLTITKELYLPDLPITRMHMLD
             IHGKRYIIDYEEINELLYKLNFTFDLIRAKLVDCANDYCOIIPNLKITANELDYLK
             KLVGRKPLDNTKDNVGNKMDYITKKNKKTIENTINLESKTKTIDKNKATYEEBK
             KLVQAOGLDLSYINQLEEAHNLISYELRIQDTLKNENITELLDKINIKNPANGG
             NTPLDLDKNRKIEHEKEIKELAKTIFKNIDSLVDPLELYYLARKNNNIDISAV
             ETKESTPEDEYPNQVYPLSYINDNNALNNLNGCLINPDYETLPEPSKNIYDNEK
             KFTIEKIKIEIKIEIESDKSVESDKSLINTDTKEYEKLENIYDSKRNNDNLTN

```


FEKMGKRYSEYKVEKLTTHNPFASYENSKHNLEKTLKALKYMEDYSLRNIIVEKELY
YKNLSIKIENEIETLVENKDEEOLFEEKITKDNENKPEKILEVSDIYKVOVOYVL
MKILDEKKITOLILNVELKHNHIVNSKOEKNPEPYLIYLKEKIDKLFEMKE
SLINEKKNIITEGOSDENSEPTEGETITQOATTKGQOAGSLBEDSVQAOKEOKA
QPEVPEVPEAKAOVPTPEAPVNPKNTEVSKIDYLEKILEFINTSTICKYILVSHST
MNEKILKOKITKEEBESKLSGCDPLDLFNIONNIPIVMSDMSLNNSLSQFMETYE
KMNVLNLYKLDNDKIKNLDEAKRVSTSVKTLSSSSMQPLSLPTQDKREVSANDTS
HSTNLNLSKLEENILSKNENKIYOELKSNSEFYEKILKSDSTGFENEFJNEFVVK
SKADDINSLEDSKRRKLEEDINKLKTOLSPDLYNKYKLERLFPDKKIVGYKM
OIKTLILKEOLESKLNSINPKHYLONSPFNKKKAEIETENTUTLENKILILKHY
KCLVITYNGESSPLKTLSESTIOTEDNTASLENFVLSLEKLDNLNLEKKISLYL
SSGLHLHLAELEKVIKKNKYTNGSNSENNVDVNMALIESYKFLPSTQDAIVDSVGS
DLLEOSOPKPASTHVGAESNTITTSQNVDDVEDVYIPIFGESEDEDYDGOVVTG
EAVTPSVINILSKIENEYEVLYKPLAGVRSKLENNMTFVNVKDIILNRFN
KRENNKLVLESDLIPIKDLTSNVYKIDPKFLNKKRDKFLSSVNYIKDSIDTINF
ANDVIGYKILSEKKSDLSIKKYINKOGENEKYPLPLNNIFLYIKRYVNDKIDLFV
IHLKAVLNTYEKSNVEYKIKELNYLKTIDKLDKFNKNNFVGIAJLSDTSTYNNHL
LTKFLSTGAVFENLAKTYLSNLDNLOGLMIISOHQVYKQCCPONSGCFRLDBREE
CCKLNTYKOGEDKVEENPPTCENNNNGGDADAKCTEEDSGSNGKKTCECTKPPSPY
LEFGIFCSSSNFLGISFLDILMLIYSFI"
1649.6508
/product="gp195 surface antigen (AA 1 - 1620)"

BASE COUNT 3050 a 727 c 803 g 2458 t
ORIGIN

Query Match 100.0%; Score 4920; DB 96; Length 7038;
Best Local Similarity 100.0%; Pred. No. 0; Indels 0; Gaps 0;
Matches 4920; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaagatcatatctcttltatgttcaatctcttltatlaataaacaagtgtga 60
DB 1592 ATGAAGATCATATCTTTTATGTTCATCTTTTATTAATAATACACATGTGA 1651
QY 61 acccagaaagtatacaagactgtcaaaaaactagaagctttagaagtgcaattg 120
DB 1652 ACACATGAAGATTATCAAGAACTTGTCAAAACCTTAGAAGTAGAGATTG 1711
QY 121 acaggtatagttatattcaaaagaaataatgtatlaaataagaagaagtgtga 180
DB 1712 ACAGGTTATAGTTATTTCAAAAGGAAAAATGTATTAATAGAGACACAGTGA 1771
QY 181 gctgttacaactagtaacacctgtgtcaaaaggttcagttgctcagggtgtca 240
DB 1772 GCTGTACACTAGTACACCTGGTTCAAAAGGTTCAAGTGTCTTCAGGTGCTG 1831
QY 241 tcaagttgctcagggtgtgtcagttgctcagttgctcagttgctcagttg 300
DB 1832 TCAGTTGCTTCAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1891
QY 301 gctcagggtgttcaggtaattcaagaagtaaaccttcagataataatcagttga 360
DB 1892 GCTTCAGGTGCTTCAGGTAAATTCACAGACGTACAAATCTTCAGTAAATTC 1951
QY 361 gatgtcaaatcttaagctgtatlaaaacacagagtaagaaatcttgttaactl 420
DB 1952 GATGCTAAATCTTACGCTGATTTAAAAACAGAGTAGAATTAAGTTAACT 2011
QY 421 gaactcaaatatcccaactcttgatttaactacatagttaacttggtgata 480
DB 2012 GAACCTAAATATCTCAACCTTTGATTTAACTAATCATATGTTAACTTTG 2071
QY 481 attcaggtttcaaatatttaattgtatgaagaataatgaattatattataa 540
DB 2072 ATTCAAGGTTCAAAATTAATTAATGATGATGAAGAAATTAATTAATTA 2131
QY 541 ttaaatcttattctgaattatlaagaagcaaaatlaaatatgtatgtgataat 600
DB 2132 TTAAACTTTTATTTGATTTATTAAGACAAATTAATATGATGATGATTA 2191
QY 601 tgtcaaataccttcaacttaaatctgtgcaaatgaattagaagtlacttaaa 660
DB 2192 TGTCAATACCTTTCAATCTTAATAATTGCGCAATGAATTAAGCTTAATAA 2251

QY 661 gtttcggatatagaacaccatlagacaatatlaaagataatgtagaagaaatg 720
DB 2252 GTTTCGGATATAGAAACCATTTAGCAATATTAAAGATATAGTAGAAATG 2311
QY 721 taacttaaaaaataaaaaaacatagaaaatataatgaattaaatgtagaag 780
DB 2312 TACATTAATAAAAAATTAATAAACCATAGAAAATATTAATGAATTAATGA 2371
QY 781 aaacaaattgaataaaataagaatgtcaactaaagaagaagaaaaaattata 840
DB 2372 AAAACATTTGATTAATAATTAAGATGCACTTAAGAAGAAAGAAAAAATTA 2431
QY 841 gttcaatagatcttctatttaacaatacaatagaagaagcacataatlaa 900
DB 2432 GCTCAATATGATCTTCTATTATTAATTAACATTAAGAAAGACATATTA 2491
QY 901 gtttagaaaaacgtatgacacttlaaaaaaaatgaaacatlaaggaatt 960
DB 2492 GTTTTGAAGAAACGTATTTGACACTTTAAAAAATAAGAAACATTAAG 2551
QY 961 aagataaatgaattaaaaatcccccacccgccaattctggaatatcacc 1020
DB 2552 AAGATTAATGAATTAATAATCCCAACCGCCCAATTCGAAATACACCAAT 2611
QY 1021 ctgtataagaacaaaaaatcgaagacacgaagaaataaagaatgccaacct 1080
DB 2612 CTGTATATAGACAAAATAATCGAGACACGAAACGAATAATAAATAATTC 2671
QY 1081 attaaatttaattatgatttatttactgacacttgaattgaattactat 1140
DB 2672 ATTAATTTAATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 2731
QY 1141 gaaaaaataaaaaattatataagtgcaaaagttgaaacaaagaatacaact 1200
DB 2732 GAAAAAATAATAAATATGTATATAGTGCAAGAGTTGAACAAAGAAATCA 2791
QY 1201 aatgaatcccaatgagttacttaccctgtgtcatalaagataatgaatg 1260
DB 2792 AATGATATTCGAATGAGAGTACTTTCCTTGTCTATATAGCATTTAA 2851
QY 1261 aatgaacttaattcttctgtgtgtatlaatcaatcattgtatatacaaa 1320
DB 2852 AATGAACCTTAATCTTTGCTGATTTAAATTAATTAATTAATTAATTA 2911
QY 1321 aaaaatatatatctatataatgaagaaaaaatctataatgaataatga 1380
DB 2912 AAAACATATATATCTGATTAATGAAGAAAAAATTAATTAATTAATTA 2971
QY 1381 aaaaatgaagaaaaaaatltgaatcgtatlaaaaaatcttcaagaagac 1440
DB 2972 AAATATGAAGAAAAAATAATGATCTGATTAATAATTTACGAACACGAT 3031
QY 1441 ttaaatgalataacaaagaataltgaanaattacttaatgaattatga 1500
DB 3032 TTAATATGATTAACAAAGAAATATGAATAATTAATTAATTAATTAAT 3091
QY 1501 aataataatataagatttaactaaatttcgnaaaaaatgaatggtaaaaga 1560
DB 3092 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3151
QY 1561 gttgagaacttaacacacataacttctgcatcctatgaagaatctaa 1620
DB 3152 GTTGAGAAACTTACACACATTAATTAATTAATTAATTAATTAATTAAT 3211
QY 1621 gaaagttaacaaagacttcaaatatagaagaatcttcttlaagaaatag 1680
DB 3212 GAAAGTTAAACAAAGCTCTTAATTAATTAATGAAGATTTATTTAGGAAT 3271
QY 1681 gaaagaatataaatatataaanaatttaataagcaaaatgaagaatga 1740
DB 3272 GAAAGAAATTAATAATTAATTAATAATTAATTAATTAATTAATTAATTA 3331

1741 ttagtctgaataatctaaagaatgaagaacagcttcttgaaaaaaatctactaaagac 1800
|||||
3332 ttatgtgaaatatttataaagaatgaagaacaccttttgaaaaaaatattactaaagac 3391
Db
1801 gaaataaaccagatgaagaatctttagaagtactgcacatgtgtaaaagtaacgttcaa 1860
|||||
3392 gaaaaaataacacagatgaaaaaatttttagaagatctgcacattgttaaaagtaacagttcaa 3451
Db
1861 aaagttcttaataagcaaaatctgcaaatcttaaaagaacatcgaattcttaaaaaat 1920
|||||
3452 aaagttttattatgaacaaaattgacgaatttaaaagaactcaattgatttttaaaaaat 3511
Db
1921 gtagaattaaacatgatacactgtcccaattcttacaacaaagaataaagaacagaa 1980
|||||
3512 gtagaatttaaacatgataatgactgttcccaattcttacaacaaagaataaagaacagaa 3571
Db
1981 ccttactatctaatgtgtctgtaaaaaagaatctgataaattaaagtgctcagctaaag 2040
|||||
3572 ccttattatttttaattgtgtgaaaaaagaattgataaatttaaaagtgctcagctaaag 3631
Db
2041 gtagaattctatgataatgagaagaataaagaataaagaacgaagtgtaactgcgataat 2100
|||||
3632 gtagaattctattgtatgaataaagaataaagaataaagaacgaagtgtaactgcgataat 3691
Db
2101 tcggaaccatacaaccgaaggaagaatacaacaggaacgaactacaacacgtggaacaa 2160
|||||
3692 tcggaaccatacaaccgaaggaagaatacaacaggaacgaactacaacacgtggaacaa 3751
Db
2161 gcaagactgtcctttagaaggaagatctcagtaacaagcaacgaacaaagaacaaacagca 2220
|||||
3752 gcagagactgtcctttagaaggaagatctcagtaacaagcaacgaacaaagaacaaacagca 3811
Db
2221 caaccacccagttccagttaccaggaacgaagaagcaaaagcaagtcaccaacacacagca 2280
|||||
3812 caaccacccagttccagttaccaggaacgaagaagcaaaagcaagtcaccaacacacagca 3871
Db
2281 ccagtaataataaactgaagaatgtcttcaaaatagatctcttgaataatctataatga 2340
|||||
3872 ccagtaataataaactgaagaatgtcttcaaaatagatctcttgaataatctataatga 3931
Db
2341 tttttaataactctcatatataatgtaacaaatataatcttggttccacacacacatgaac 2400
|||||
3932 ttttttaaaactctcatatataatgtaacaaatataatcttggttccacacacacatgaac 3991
Db
2401 gaaaataataaataaataaataaataaataaagaaggaagaagaacgaatgaatctatgt 2460
|||||
3992 gaaaataataaataaataaataaataaataaagaaggaagaagaacgaatgaatctatgt 4051
Db
2461 gatccatagactctatgttcttaataatacaaaataacatacctgttaatgtctatgtct 2520
|||||
4052 gatccatagactctatgttcttaataatacaaaataaataacatacctgttaatgtctatgtct 4111
Db
2521 gatagtcttaacaataatgttcttaatacaacatacttaagaaatttaagaaagaatgtgt 2580
|||||
4112 gatagcttttaaacatgattgttatcacaacattttatgaaatttttaaaaaaagaatgtgt 4171
Db
2581 tattaatttaataaacttaagaataatgaagaataaataaataaataatcttaagaagaagcaaa 2640
|||||
4172 tattaattttatataaacttaagaataatgaagaataaataaataaataatcttaagaagaagcaaa 4231
Db
2641 aaagtaatacaacatctgtaaaaaactcttcaagtctcaacgaacacatactatcaatcaaa 2700
|||||
4232 aaagtaatacaacatctgtaaaaaactcttcaagtctcaacgaacacatactatcaatcaaa 4291
Db
2701 cctcaggaataaaccggaagtaagtgcaaatgataatgatacacaatctcaaaattgaaat 2760
|||||
4292 cctcaggaataaaccggaagtaagtgcaaatgataatgatacacaatctcaaaattgaaat 4351
Db
2761 aataagtttaaaatcttgaagaacatactgagctcttggaagaagaagaataatatacaaa 2820
|||||
4352 aataagtttttaaaatcttgaagaacatactgagctcttggaagaagaagaataatatacaaa 4411
Db
2821 gaattaataagtcataaaagaatgagtgaaacattctgaagaagatactaaagaatgagat 2880
|||||

4412 gaattatgactgcaaaaaagtaagtaaaacttttatgaaaagatttttaaaagtaactgact 4471
Db
2881 acattctataatgaatccttcttaacaattctgttaaaatctaaagctgaatgatatctca 2940
|||||
4472 acatttttatataaattcttttacaaattttgttgaatttttaaaactgataatattatctca 4531
Db
2941 ttagatgataatcaaaagaagaagaatttagaagaagatactaaatctaaaaaaaac 3000
|||||
4532 ttgaatgataatcaaaagaagaagaatttagaagaagatactaaatctaaaaaaaac 3060
Db
3001 ttaacatctatcatcttgattatataataataatactaaatctagaagaattatctgact 3060
|||||
4592 ttacactgttatcatcttgattatataataataatactaaatctagaagaattatctgact 4651
Db
3061 aaaaagaataaaccagtggttaataataaataatgaagaataaanaactctatataaaga 3120
|||||
4652 aaaaagaataaaccagtggttaataataaataatgaagaataaanaactctatataaaga 4711
Db
3121 caattagatcaaaatctgaattctacttaaacccaagaagcatctatacaaaactcttc 3180
|||||
4712 caattagatcaaaatctgaattctacttaaacccaagaagcatctatacaaaactcttc 4771
Db
3181 gttctcttaacaataaagaagaagctggaatagcaagaactgaagaacacatctagaagaac 3240
|||||
4772 gttttcttttaacataaagaagaagctggaatagcaagaactgaagaacacatctagaagaac 4831
Db
3241 acaaaaatactatgaagaacatataaagaagctgtttaaatalatataatggtgaatcatct 3300
|||||
4832 acaaaaatactatgaagaacatataaagaagctgtttaaatalatataatggtgaatcatct 4891
Db
3301 ccaattaaaactttagatgaagaatcaatctaaacgaagaataatctagccagtttagaa 3360
|||||
4892 ccaattaaaactttagatgaagaatcaatctaaacgaagaataatctagccagtttagaa 3420
Db
3361 aactttaaagatctaaagtaaatctagaagaagaataaagaatctaaatctagaagaag 3480
|||||
4952 aactttaaagatctaaagtaaatctagaagaagaataaagaatctaaatctagaagaag 5011
Db
3421 aaaaatactatcatcttaacaaatctgaagaatcaatcttaatctgtgaatataaagaagta 3480
|||||
5012 aaaaatactatcatcttaacaaatctgaagaatcaatcttaatctgtgaatataaagaagta 5071
Db
3481 ataaaaataataaataatctaaagtaatcttcaagaagaagaataatctgaagaatctaa 3540
|||||
5072 ataaaaataataaataatctaaagtaatcttcaagaagaagaataatctgaagaatctaa 5131
Db
3541 gcatlagaatcttacaataaactctcccaagaagaagaatctgtcaacagctgtgaagt 3600
|||||
5132 gcatlagaatcttacaataaactctcccaagaagaagaatctgtcaacagctgtgaagt 5191
Db
3601 gaaagtgtgactcgaacacatctagaagaagaatctcaacaaagaacgaacacatcaatctga 3660
|||||
5192 gaaagtgtgactcgaacacatctagaagaagaatctcaacaaagaacgaacacatcaatctga 5251
Db
3661 gtagcagagctcttaacaaatacaacatacaaaatgtctgaatgataatgataatgacgtga 3720
|||||
5252 gtagcagagctcttaacaaatacaacatacaaaatgtctgaatgataatgataatgacgtga 5311
Db
3721 atcatagtaactctatcttgagaatctcgaagaagaatctgaatgataatctagaagaagta 3780
|||||
5312 atcatagtaactctatcttgagaatctcgaagaagaatctgaatgataatctagaagaagta 5371
Db
3781 acagagaagaacgaatcaactccttcgtaattctgaatacaatctcttaaaatctgaagaatga 3840
|||||
5372 acagagaagaacgaatcaactccttcgtaattctgaatacaatctcttaaaatctgaagaatga 5431
Db
3841 tatgaggtttatatttaaaactcttagcagagtggttataagaagtttaaaaaacata 3900
|||||
5432 tatgaggtttatatttaaaactcttagcagagtggttataagaagtttaaaaaacata 5491
Db
3901 gaaaataacagctatgacatctaatgttaagtataatcttaaaatcacagatctaat 3960
|||||

Db 5492 GAAATATACGTTATGACATTTAAATGTTAAATGTTAAAGATATATTTAAATCAGCATTTAAAT 5551
 Qy 3961 aaacgtgaacattccaaaaatgttttagaatccagatttaattccataataagaattaca 4020
 Db 5552 AAACGTGAACATTTCAAAATGTTTATGATTCAGATTATTCATATTAAGATTTTACA 5611
 Qy 4021 tcaagtaataatgttgcacaaagatccataaaattcttaataaagaanaagagataaa 4080
 Db 5612 TCAGATATATATGTTGTCACAAAGATCCATATAATTTCTTAATAAGAAAAGAGATAAA 5671
 Qy 4081 tttcttaagcgtttatattatattaaagattcaatagatccagatataaaatttgcgaat 4140
 Db 5672 TTTCTTAAGCGATTATATATATTAAGATTCATATAGATTAATAATTTTGGCAAT 5731
 Qy 4141 gacttccttgatattataaataattatccgaaaaataataatcagaattcagaattcaatc 4200
 Db 5732 GATGTTCTTGATATATATTAATAATTTATTCGAAAAATATAATACAGATTTCAGATTCAAT 5791
 Qy 4201 aaaaaatatacgaacgaacaaagtgaaaaatgagaataaccctcccttttaacaat 4260
 Db 5792 AAAAAATATATCAACGACAAACAGTGAAATGAGAAATACCTCCCTTTTAAACAAAT 5851
 Qy 4261 attgagaacctatataaaagaatgaatgaataattgattatttgaattcaattaga 4320
 Db 5852 ATTGAGACCTTATATTAATAAGATTAATAATGATTAATAATGATTAATTTGATTAAGAA 5911
 Qy 4321 gcaaaagcttcaataataacatacgaacaaacgaagtagaagttaaataaagaacatc 4380
 Db 5912 GCAAAAGTTTAAATTTATATATGAAATTCAGAACTTACAAAGTTAAATTAAGAACTT 5971
 Qy 4381 aattacttaaaaaaatccaatccaagaacaaattggcagattttaaaaaataaacaattcgt 4440
 Db 5972 AATTACTTTAAAAACAATTCAGACAAATTCGAGATTTTAAAAAATAACAATTCGTT 6031
 Qy 4441 ggaattgctgattatcaacaaagattataacataataactatattgacaaagttccatg 4500
 Db 6032 GGAAATGCTGATTTATCAACAGATTAATAACCAATTAATACTTATTTGACAAAGTTCCCTAGT 6091
 Qy 4501 acagatgattgttttgaacatctgtcctaaacgctttatctaatcttactgtagtgaaac 4560
 Db 6092 ACAGATATGTTTGTGAAATCTTCTGCTAAACCGTTTATCTTAATTTACTTGATGGAAC 6151
 Qy 4561 ttgcaagatgtttaaacaattccacacacaaatgctgtaaaaaaacaatgtccacaacaaat 4620
 Db 6152 TTGCAAGATGTTGTTAAACATTTTCAACACCAATGCGTAAAAAAACAATGTCCACAAAT 6211
 Qy 4621 tctgagatgtttcagacatttgaagaagaagaatgttaatttattataataattaca 4680
 Db 6212 TCTGATGTTTTCAGACATTTTGAATGAAGAAGAAATGTAATGTTTATTAATTTACAAA 6271
 Qy 4681 caagaagtgatataatcgtgttgaacaaatccactctgtaacgaaaaataatgttga 4740
 Db 6272 CAAGAAGTGATATATCTGTTGAAATCCAAATCTTACTTGTAAACGAAATTAATGTTGGA 6331
 Qy 4741 tgtgatgcagatgacaaatgttacgaagaagaatccagtagcaacggaagaagaatcaca 4800
 Db 6332 TGTGATGCGAGATGCCAAATGTACCGAAGAAGATTACAGTGCACAAAGAAAGAAATCACA 6391
 Qy 4801 tctggaatgtactaaacctgattcttaccacttccgattgtatttctgcaattccctc 4860
 Db 6392 TGTGAATGTACTAAACCTGATTTTATCCACTTTTCGATGATTTTCTGCACTTCCTCT 6451
 Qy 4861 aactttcttgaataatcattcttattataactatctatattataatcagttcatttca 4920
 Db 6452 AACTTCTTAGAATATCATTTCTTATTAATTAATCATGTTAATTAATACAGTTTCATTTAA 6511

RESULT 4
 LOCUS A04562 5760 bp DNA PAT 13-JUL-1993
 DEFINITION P.falciparum P.195 gene.
 ACCESSION A04562
 VERSION A04562.1 GI:410754

KEYWORDS malaria parasite P. falciparum.
 SOURCE Plasmodium falciparum
 ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 5760)
 AUTHORS Holder, A.A., Sandhu, J.S., Odink, K.G., Lockyer, M.J. and Riveros-Moreno, V.
 TITLE Cloning of DNA for protozoal antigens
 JOURNAL Patent: EP 0154454-A 2 11-SEP-1985;
 THE WELLCOME FOUNDATION LIMITED
 FEATURES
 source location/Qualifiers
 1..5760
 /organism="Plasmodium falciparum"
 /db_xref="taxon:5833"
 216..5180
 /gene="P.195"
 216..5180
 /gene="P.195"
 /codon_start=1
 /protein_id="CAA00367.1"
 /db_xref="GI:410755"
 /translation="MKIIPIICSPLEFIITNOCYTHESVOELVKTLEADVLVLYGS
 LFOKERMTNBECSGTAATTTTPGSKGASGSGSVASGSGSVASGSGSVASGSGSVAS
 GSGNSRRTNPSDSSSDAKSYADLKHRYNRYLLITELKYPOLDELTHNMLTLCDN
 HGPFYLDIGYEINELLYKLFNFPDLAKLNDYCANDCOIPNLKIRAMELDVLK
 KLVGFRKPLDNIKDNVGMEDYIKKKNTIENINELIEESKTTIDKKNAREEEK
 KLYOAYDLSTYNKOLEEHNLSVLEKRIIDPLKNNENIKELDKINEIKNPPANG
 NPNTLDKNNKIEIEHEKRIETATIKRINDISLPDLELEYIAREKKNINDISKV
 ETKESTEPENYNGVYPLSYNDINNALENSFGDLNFPDYTKEPSKNITTDKER
 KFINETKRIKLEKKIESDKSYEDRSLSNDITREKLELNIEYDSKFNNTIDTN
 PERKMGKRYSVYERKTHPNTFAYSYNSHNLEKLEKALYKVEDSLRNIVYKELKY
 YKNLSIKLENEIEKLEKDEDEFEKIKRDNKPEKILEVDYIVVOYVLL
 MNKIDELKKTOLILKNVELKHNHIVNSYKONKOEPPYLLIYKKEIDKLEKFMKVE
 SLINNEKNIKITOGOSDNSEPTSEETICGATKCGOAGSLBEDSOAOROKOA
 OPPVYPPPEAKAQQPTPPAPVNNKTENVSKIDYLEKLOPLNTSTICHKTLYVSHST
 MNEKILKOYRTKEESKLSGCDPLDLFNIONLIPVMISPDSTINIVHNLWYER
 EIGVYVILMETYERKEMVCNLKYLKDNKIKMLLEAKVSVSTVLTSSSSMOPVLT
 PODKPEVSANDTSHSTNLNLSIKLEFNELISGKNKIYQELIGKSSNFYEKILKD
 SDTPFNSEFTNPKSKADINDSLNLESKSKKLEEDINIKTKTLOLSPDLVNNYKLE
 RLFDKKRYVKKMOIKITLIKEDOLESLNLPKHYLONFYFPKKRAELET
 ENTLNERTKLLKHYGLVAYNGESSPLETLEESIIYEDNTASLENREKRAELET
 KDNLMLEKKILSYLSRGLHHLIAELKEVLTKNYKNYSVNNTDVNNALESYKFLP
 ECTDVAIVVSESGSDTLDOOPKRPASTHVGAESNTITTSQWDEVDVIVLTFGE
 SEEDYDLDGVVYTGAVTTSVIDNLISIKLENEYLYLKLPGVYRSLKOLENNMT
 FNVNKKDILNSRFNKRNFKNVLESDLPYKDLTSNVYKDPYKFLNKKERDKFLSS
 VNYIKDSIDPDINDPANDVIGYKTIISSEKSDSITIKYINDKOEENKYLEPUNIE
 TIKTVNDKIDILFVHLEKAVLNTYTESNVEKIKELVLTIDKIDKLADEKKNV
 GIADISTDYNHNNLLTKFLSTGMVFNELKSVLSMLMKLARVYKHTTPMRKKTMT
 QOSSCFRHLDERBECKLNLNKOEGSKVENSNPFCNENNGGCDADKCTEEDSGST
 GKRTICECTKPCYPLSMWIFCSSNFELGISPLILIMILVYFI"
 BASE COUNT 2565 a 630 c 725 g 1840 t
 ORIGIN
 Query Match 97.1%; Score 4779.2; DB 9; Length 5760;
 Best Local Similarity 98.4%; Pred. No. 0;
 Matches 4886; Conservative 0; Mismatches 33; Indels 47; Gaps 4;

241 tcaagttgctcaagtgagtcagttgctcaagtgagtcagttgctcaagtgagtcagtt 300
|||||
456 TCAGTTGCTTCAGTGAGTGCTGCTGAGTGCTGAGTGCTGAGTGCTGAGTGCT 515
OY gcttcagtgagtcagtgatctcaagacgctacaaacccttcagatctcaagtgatct 360
|||||
516 GCTTCAGTGAGTGCTGAGTGATTCAGACGTCACAAATCCTTCAGATTAATTCAGTTCA 575
OY gatgctaaatcttaagcgtgacttaaaacagagtgacgaatctctgttaactcaaa 420
576 GATGCTAAATCTTACCGCTGATTTAAACACAGAGTACGAAATCTGTTAACTATCAAA 635
OY gaactcaaatatcccaactcttgatcttaactcaatcatatgtaactctgtgtaaat 480
636 GAACCTCAAAATCTCCTCACTCTTGAATTAATCAATATGTTAACTGTTGTTGAAT 695
OY atctcagtgctcaaatatcttaattgaatgagatgaaatlaaatatctataaa 540
696 ATTCATGCTTCAAAATATTTAATGATGATGATGAAGAATTAATGATTAATTAATTA 755
OY ttaacttctatcttgatcttaagacaaatlaaatgagtgatgcttaagtat 600
756 TTAACCTTTATTTGATTTATTAAGCAAAATTAATGATGATGATGCTAATGATTAAT 815
OY tctcaaatccttcaatcttaaatcttgacagatgagatgagcttaaaact 660
816 TGCAAAATCTTCACTCACTTAAATTCGTCGAATGAATTAACGCTACTTAAAAACTT 875
OY gcttcagtgatagaaacatctagaacatctaaagatctgaagaaatctgaagat 720
876 GCTTTCGATATTAAGAAACCATAGACATATTAAGATTAATGAGAAATTAAGAAAT 935
OY tacatctaaataaataaataaaccatagaaataataatgaaatctgaagaaatgaag 780
936 TACATTAATAAAATTAATAAAACCATAGAAATTAATAATTAATTAAGAAATGAAG 995
OY aaaaacatctgaataaataaagatctgaactaaagaaagaaataaataaataatctccaa 840
996 AAAACATTTGATTAATAATTAAGATGCACTTAAGAAAGAAAGAAAGAAATTAATCCAA 1055
OY gctcaatctgact 900
1056 GCTCAATATGATCTTCTCACTTTCATTAATAACATTAAGAAAGCACTAATTAATTAAG 1115
OY gttctgaagaaacgtatctgaactcttaaaataaataatgaacatctgaagatctgag 960
1116 GTTTTGAATAAACGTATGACACTTTAAATAAAATGAATAACATTAAGCAATTAAGTGAAT 1175
OY aagatctaaatgaatctaaataatccccaacgagcgaactctggaatctacacaaatctctc 1020
1176 AAGATTAATGAATTAATAATCCCAACGCGCAATTTGGAATTAACCAATAATCTCTC 1235
OY cttgataagacacaaataaataaataaataaataaataaataaataaataaataaataa 1080
1236 CTGTGATTAAGACAAAAATTCGAGGAACAGAAAAAGAAATTAATAAATAATTCGCAAAACT 1295
OY actaaatcttaataatgagtgatcttaactgactgacactggaatctgaatacttaaga 1140
1296 ATTAATTTAATAATGATGATGATTAATTAATTAATTAATTAATTAATTAATTAAG 1355
OY gaaataaataaataaataaataaataaataaataaataaataaataaataaataaataa 1200
1356 GAAAAAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1415
OY aatgaaatctcaaatgagtgactctctctctctctctctctctctctctctctctctct 1260
1416 AATGAATATCAAAATGAGTACTTCTCTTGTCAATTAAGATTAATTAATTAATTAATTA 1475
OY aatgaaatct 1320
1476 AATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1535

1321 aaaaacatctatactgataatgaaagaaaaatactcaataatgaatctgaagaaaaatct 1380
|||||
1536 AAAACATTAATATCTGATTAATGAAGAAAAATTTATTAATGAATTAAGAAAAATTT 1595
OY aaaaatgaataaataaataaataatgaaatctgaataaataaataaataaataaataaata 1440
1596 AAAATGAATAAATAAATAATTAATGAATCTGATTAATAAATAATTAAGAAAGATTAAGTCT 1655
OY ttaataatgataaataaataaataaataaataaataaataaataaataaataaataaataa 1500
1656 TTAATGAATTAATAAATAAAGAAATTAATAAATAATTAATAAATAATTAATTAATTAAT 1715
OY aataataataatgagtgatcttaactaaatctcgaataaataaataaataaataaataaata 1560
1716 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1775
OY gctcgaagaaactctcacaacgaataaactctcgaataaataaataaataaataaataaact 1620
1776 GTTGAGAACTTACACACGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1835
OY gaaagtgaaacaaagcctcttaataatgaaagatctctctctctctctctctctctctct 1680
1836 GAAAGCTTAACAAACCTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1895
OY gaaagaaatctaaatcttaataaataaataaataaataaataaataaataaataaataaata 1740
1896 GAAAGAAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1955
OY ttaagtgaaataaataaataaataaataaataaataaataaataaataaataaataaataa 1800
1956 TTAGTGAAATAATTAATAAAGATGAAGAACAGCTTTTGAAAAATAATTAATTAAGAAC 2015
OY gaaataaataaataaataaataaataaataaataaataaataaataaataaataaataaata 1860
2016 GAAATTAATAACCAATGAATAAATAATTAATAAATAATTAATAAATAATTAATAAATA 2075
OY aagaatcttaataaataaataaataaataaataaataaataaataaataaataaataaataa 1920
2076 AAGCTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2135
OY gtaagaatctaaataaataaataaataaataaataaataaataaataaataaataaataaata 1980
2136 GTGAATTAATAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2195
OY ccttataatcttaatctgcttgaaataaataaataaataaataaataaataaataaataaata 2040
2196 CCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2255
OY gtagaatctatgataaataaataaataaataaataaataaataaataaataaataaataaata 2100
2256 GTAGAAATCATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2215
OY tctgaaacatctcaacgaaagaaataaataaataaataaataaataaataaataaataaataa 2160
2316 TCGAAGCAATCAACGAGAGAAATTAACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 2275
OY gcaagatctgctcttagaagagatctgaataaataaataaataaataaataaataaataaata 2220
2376 GCAAGATCTGCTTTAGAGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2435
OY caacacacagtgacagtgacagtgacagtgacagtgacagtgacagtgacagtgacagtgac 2280
2436 CAACCAACGATCACTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 2495
OY ccaagtaataaataaataaataaataaataaataaataaataaataaataaataaataaataa 2340
2496 CCAAGTAATTAATAAATACTGAATAATGTTTCAATTAATTAATTAATTAATTAATTAATTA 2555
OY ttttaataactctataatgataaataaataaataaataaataaataaataaataaataaataa 2400
2556 TTTTAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2615
OY gaaagataatctaaataaataaataaataaataaataaataaataaataaataaataaataa 2460

Db	2616	GAAGAATTTAAACAAATTAATTAATTAACAAAGCAGAAAGAAAGCAATTAAGTCAAGT	2675
Qy	2461	gaccatbagactatgtgttaataacaaabaacataccctgtaaagtatctatgtt	2520
Db	2676	GATCATTAGACTTATGTGTTTAATATACAAATAATACATACCTGTATGATCTATGTGTT	2735
Qy	2521	gatagtctaaacaatagttatcacacact	2524
Db	2736	GATAGCTTAAACCTAATCTTATATACAACTATATATGCGTTATGAAAAAAGAAATTTGGTTA	2794
Qy	2552	-----tlaaggaaattatgaaagaagcggcttgtaattatataaact	2597
Db	2795	TGTAATTATATATCTTATATGCAAAATTTATGAAAAAGAAATGGTTGTATTTATATAACT	2854
Qy	2598	laaggaataatgacaaaataaaaaattatlagaggaagcgaaaaaglatccacatcgt	2657
Db	2855	TAAAGATATATGACAAATTAATTAATTAATTAAGAGCAAGCAAAAAAGATCCACATCGT	2914
Qy	2658	aaaaactcttcaagttcatcatgcaacacatataaataacactcagatlaaacccga	2717
Db	2915	AAAAACTCTTCAAGTTCATCAATGCAACCATATATCATTAACACCTCAGGATAAACCCGA	2974
Qy	2718	agtaagcgcaatgataacacacatccatccacactcacaattggaataagttlaaaatatt	2777
Db	2975	AGTAAGTGCANATGATGATATCATTCACATTCACAAATTTGAAATTAATAGTTTAAATTAAT	3034
Qy	2778	tgaaaaacatactgagctcttgaaaaaaacaaaatatalaccagaattaatagtcacaa	2837
Db	3035	TGAAACATATATGTGCTGTGAAAAACAAAAATATATGCCAAGATTAATATAGTCAAAA	3094
Qy	2838	aagtatggaacactttagaaaaagatatlaaagatagtgatatacttataatgaac	2897
Db	3095	AAGTAGGAAAACTTTATATCAAAAGATATTTAAAGATGATGATCAATTTATTAATGAATC	3154
Qy	2898	tttacaacatttgtaaaatctaaagctgaagatataatgaattgaatgaatcga	2957
Db	3155	TTTTACAAATTTGTATAAATCTTAAGCTGATGATTAATTAATGATGAATGATGAATCAAA	3214
Qy	2958	aaggaagaaattgaagaagaagatatlaataatlaaaaaaacctttacagttacattga	3017
Db	3215	AAGGAAGAAATTTAGAAAGATATTTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3274
Qy	3018	tttatataataataataaataaaltbagaaagatatattgataaaagaacacagttg	3077
Db	3275	TTTTATATAATAATTTAAATTTAAATTTAAAGATATTTGATTTAAAGAAACAGTGTG	3334
Qy	3078	taaatataaabaacaaattaaaaaacttacttatataaagaacatrtagaatcaaat	3137
Db	3335	TAAATATTAATATGCNAATTTAAAAACCTTACTTTATTAAAGAACATTTAGATTCAAATTT	3394
Qy	3138	gaattcaacttaataaccacaaagcatglatlacaaaactttctgttcttcttaacaaaa	3197
Db	3395	GAATTCACCTTAATTAACCAAGATGTATTTACAAAACCTTTCTGTTCTTTACAAAAA	3454
Qy	3198	aaaaagaagctgaaatagcagaaacggaanaacacatagaaaaacacaaaatatatgaa	3257
Db	3455	AAAAAGAAGCTGAAATATGCAAAAACCTGAAAAACATTTAGAAAAACAATAATTAATTGAA	3514
Qy	3258	acattataaaggacttgtaaatatataatgaatgigaatcatctccattaaaaaacttaag	3317
Db	3515	ACATTATTAAGAGACTTGTTAATTAATTAATTAATGATGATATCTCCATTTAAAAACTTTAAG	3574
Qy	3318	tgaaagatacaattcaaacagaagaataattatgcgaatttagaaaacttlaagataaag	3377
Db	3575	TGAAGATCATCTTCAAAACACACATATATATGTCGACTTTAGAAAACTTTAAAGTATTAG	3634
Qy	3378	taaattagaagaaaattlaaaggataaacttbaatttagaagaagaanaatatcatctt	3437
Db	3635	TAAATTTGAAGAAATTAATTAAGATTAATTTAAATTTAGAAAAAAGAAATTAATCATCTT	3694
Qy	3438	ataaagtggaattacatttaattatgctgaaattlaaagaagtaataaaaaataaaatla	3497

[illegible]

Oy	4577	acatttcacaacacccaatctgcgtataaaaaacaatg- Tpaccaaaattctcgtatglttcag	4634
Db	4835	ACATTTTCAACAACACCAATGCGGTAAAAAACAATGATGCACACACAGTCTGGATGTTTCAG	4894
Oy	4635	acatttgaatgaaaagagaagaatgtcaatglttataatatacaacaagaagtgatataa	4694
Db	4895	ACATTTTGAATGAAAAGACAAAGAAATGTAAAGTCTTATTTAAATTTCAAAACAAGAAGGTAGTAA	4954
Oy	4695	atgtgttgaaaatcccaatctcactctgttaagaaaaatagtgltgatgtgagatgac	4754
Db	4955	ATGTGTTGAAAATTTCAAAATCTTACTGTAAACGAAATTAATGCTGATGTGATGCAGATGC	5014
Oy	4755	caaatgtacccaagaagaatccagtagacaagaaagaanaatcacatgtgaatgtactaa	4814
Db	5015	CAATTTTACCGAAGAAAGATTCAGGTGACACGAAAGAAATACATGTGAATGTACTTAA	5074
Oy	4815	accatgaattctatccacatttgcgaatggtatattctgcagttctcttaacttcttagaat	4874
Db	5075	ACCTGATGTTTATTCACACTTTTGATGTGTAATTTTCTGCATCTCTCAACTTCTTTAGAAAT	5134
Oy	4875	atacttcttatataactcctgtataatataatacagtttcatltaa	4920
Db	5135	ATCATTTCTTATTAATACTCACTGTTAATATTATTAACAATTTCATTTTAA	5180

RESULT	5	
E00656		
LOCUS	E00656	5760 bp RNA
DEFINITION	cDNA encoding p.195 protein of plasmodium falciparum.	
ACCESSION	E00656	
VERSION	E00656.1	GI:2168935
KEYWORDS	JP 1986019490-A/1.	
SOURCE	JP 1986019490-A/1.	
ORGANISM	Plasmodium falciparum.	
REFERENCE	Plasmodium falciparum	
AUTHORS	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
	1 (bases 1 to 5760)	
	Ansoni,A.H., Malkuru,J.R., Jiyasubida,S.S., Barentelina,R.M. and	
	Kareu,G.O.	
TITLE	DNA CLONING FOR ANTIGEN OF PROTOZOA	
JOURNAL	Patent: JP 1986019490-A 1 28-JAN-1986;	

TITLE	JOURNAL	COMMENT
DNA CLONING FOR ANTIGEN OF PROTOZOA	Patent: JP 1986019490-A 1 28-JAN-1986;	
WEILCOMÉ FOUND LTD:THE	OS Plasmodium falciparum	
PN JP 1986019490-A/1		
PD 28-JAN-1986		
PF 22-FEB-1985 JP 1985034280		
PR 22-FEB-1984 GB 84 8404692, 26-SEP-1984 GB 84 8424340 PI		
ANSONIT AASAA HORUDAA, MAIKIRU JIEMOSU RONSUKIYVAA, PI		
JITASBIRAA SHINGU SANDEVU, BARENTEINA RIBEROSU MORENO, PI		
KAREER		
GERITSUTU ODEIKURU		
PC C12N15/00,A61K39/015,C07H21/04,C07K13/00,C12N1/20,C12P21/02,		
PC (C12N1/20,		
PC C12R1:19),(C12P21/02,C12R1:19);		
CC strandedness: Single;		
CC topology: Linear;		
CC hypothetical: No;		
CC anti-sense: No;		
Key	Location/Qualifiers	
PH		
PH		
FT CDS	216..5180	
FT /product='P.195 protein'		
FT sig_peptide	219..272	
FT mat_peptide	273..5177	
FT /product='P.195 protein'.		

FEATURES	/Protein-11.75 protein
source	Location/Qualifiers 1..5760
BASE COUNT	/organism="Plasmodium falciparum" /db_xref="taxon:5833"
ORIGIN	/db_xref="taxon:5833"
	2565 a 630 c 725 g 1840 t

Query Match 97.18; Score 4779.2; DB 10; Length 5760;

Best Local Similarity	98.48;	Pred. No. 0;
Matches 4886; Conservative	0;	Mismatches 33; Indels 47; Gaps 4;

QY	1	atgaagatcatatctcttttaigtatctctctttttatataataatacaaatgta	60
Db	216	atgaagatcatatctcttttaigtatctctctttttatataataatacaaatgta	275
QY	61	aaacatgaacgttcaagaacctgtcaaaaaactagaaccttgaagaatgcagtctg	120
Db	276	ACACATGGAAGTTTCAAGAACTTGCAAAAACTTGAAAGCTTTAGAAATGCACTATTG	335
QY	121	aaaggtatagttatcttcaaaaagaanaaagtgtatlaaatgaagaaacaagtgaaca	180
Db	336	ACAGGTTATAGTTATTCTTCAAAAAGGAAAAAATGTTAAATGAAGAAACAGTGAACA	395
QY	181	gctgttaaacatctagaacctggttccaaggggttcaaggttgcttcaaggttgctcag	240
Db	396	GCTGTTAACAATCTATPACACCTGGTTCAAAAGGTTCCAGTTGCTTCAGGTGGC	455
QY	241	tcaagttgttcaaggctgcacagttgtctcgaagtgagctcagttgcctcagtgctc	300
Db	456	TCAGTTGCTTCAGGTGGCTCAGTTGCTTCAGGTGGCTCAGTTGCTTCAGGTGGCTC	515
QY	301	gcttcaaggttggttcaaggtlaattcaagaacgtacaaatcccttcagataaatcaagtgatca	360
Db	516	GCTTCAGGTGGTTCAGGTAATTCACAGCAGTACAAATCCTTCAGTAATTCAGTATCA	575
QY	361	gatgtcaaatcttcaagcttgattttaaacaagaagtagcaaatcttctgttaactataa	420
Db	576	GATGTCAAAATCTTCGGTGTGATTTAAACACAGAGTACGAATTAATCTGTAACTCTCAAA	635
QY	421	gaacctcaaatctctcaaaccttgatttcaacatcatgttcaacttggtygataat	480
Db	636	GAACTCAAAATCTCTCAACCTTTGATTACTAATCATATGTTAACCTTGGTATAT	695
QY	481	atcattggtttcaaaatlaaatatgataatgaagaatlaatgaatlaatalataaa	540
Db	696	ATTCATGCTTTCAAATPMTTAAATGATGATGAGAAATTAAGTAATTAATTAATAA	755
QY	541	ttaaaccttlaatttgatttataatgaagcgaataataatgaatgctgctaaatcat	600
Db	756	TTAAACTTTTATTTTGGATTTATTAAGAGCAAAATTAATGAATGATGCTCATATATAT	815
QY	601	tgtaaatacccttcaatctttaaattcgcgaaatgaaatgaatgaacttcaaaaactt	660
Db	816	TGTCAAAATACCTTTCAATCTTTAAATTCGTGCAAAATGAATTTAGACCTTAAAAA	875
QY	661	gtgttcgagataagaanaacatagaacaatatlaaagaataatgtcgaanaaacygaagat	720
Db	876	GTTGTCGGATATAGAAAAACCTTAGACAATATTTAAAGATATATGTGAGAAAAATGCA	935
QY	721	tacatlaaaaaaaataaaaaaccatagaanaataataatgaatlaattgaagaagtga	780
Db	936	TACATTTAAAAAAATTAATAAAACCATGTGAAAAATTAATTAATTAAGAAAGTAAG	995
QY	781	aaaacaattgttaaaaaataagaatgcgaacaaagaagaagaanaaaaaactatacca	840
Db	996	AAAACAATTTGTTAAAAATTAAGATGCACTTAAGAAAGAAAGAAAAAATTAATACAA	1055
QY	841	gtccaatatgtatcttctatctlaatacaataatagaagaacacataatlaataagc	900
Db	1056	GCTCAATATGATCTTTCTTATTTACAATTAACATTTGAAAGAACATTAATTAATTAAGC	1115
QY	901	gttttagaaaaaagctatgcaccttlaaaaaaaatgaacaactaaggaattactgcat	960
Db	1116	GTTTGTAGAAAACTATTGTGACCTTTAAAAAAGAAACATTAAGCAATTAATTAATTAAT	1175
QY	961	aagatlaaatgaanaataaaatccccaacggccaactttcggaaatagaccaaatcttc	1020
Db	1176	AAGATTAATGAATTAATAAATCCCAACGGCCAAATTCGTGAAATACACCAATTAATCTCTC	1235
QY	1021	cttgaataagaacaaaaaatcgaagaacgaanaagaataaagaatgtccaaact	1080

```
Db 1236 CTTGATAGAACAAAAATCGAGAACACGAAAAAGAAATTAAGAAATTCGCCAAACT 1295
QY 1081 attaaatttaattatgatagttattactgacccactgaattagaatctttaaga 1140
Db 1296 ATTAAATTATATATGATCTTTATTTAGTCGATCCACTGAATTAAGATCTATTTAAGA 1355
QY 1141 gaaaaataataaataatgataaagtgcaaaaggttgaaacaaaggaatcaatgaccc 1200
Db 1356 GAAAAAATTAATATATGATATATAGTCGAAAGGTTGAACAAAGGATCACTGAACCC 1415
QY 1201 aatgaataccaatgagagttacttactcttgccataaagaatlaaagatgcttta 1260
Db 1416 AATGAATATCCAAATGAGAGTACTTATCCCTTTCATATACGATTTAAATTCCTTTA 1475
QY 1261 aatgaacttaattctcttggttgatctaattacatccattgatatcaaaagaaga 1320
Db 1476 AATGAATCTTATTTCTTTGGTGTATTAATTAATTCATTTGATTAATCCAAAGAACAGT 1535
QY 1321 aaaaacataatctgataatgaaagaaaaaattcataatgaattaaagaaaaaatt 1380
Db 1536 AAAAATATATATCTGATATGAAAGAAAAAATTCATTAATTAAGAAAAAATTT 1595
QY 1381 aaaaatgaaaaaaaataatgaaatcgtataaaaaatcttaagaagacagatcgaagt 1440
Db 1596 AAAATGAAAAAATAATGAAATCTGATTAATTAATTAATTAATTAATTAATTAATTA 1655
QY 1441 ttaaatgataatacaaaagaataatgaaaaaattacttaataaattatgaatgaatc 1500
Db 1656 TTAATGATATATCAAAAGAAATATGAAAAATTAATTAATTAATTAATTAATTAATTA 1715
QY 1501 aataataataatgatttaactaatctcgaaaaaaatgagtggtaaaagataatcataaa 1560
Db 1716 AATAAATAATATATGATTTAACTAATTTGAAAAAATGATGGCTAAAGATTTATCAATFAA 1775
QY 1561 gtggagaactctcacacataactcttgccatccctatgaaaaatcttaaatctct 1620
Db 1776 GTTGAACAACTTACACCTTAATTTGATCTTGTGATGAAAAATCTTAATCAATTTCTT 1835
QY 1621 gaaaatctacaaaagctcttaataatataatgaaagattatctcttaagaaatagatc 1880
Db 1836 GAAAAATTAACAAACCTCTTAATATATGGAAGATTTATCTTAAGGAATATATAGATT 1895
QY 1681 gaaaagaattaaataatctataaaaaatttaataagcaaaatagaataatgaaatga 1740
Db 1896 GAAAAAGATTTAAATATTTATTAATAATTTAATTAAGCAAAATGAAATGAGATTGAAC 1955
QY 1741 ttaagtgaataataatataaaaaagatgaaagacgcttttgaaaaaaaattactaaag 1800
Db 1956 TTAGTTGAAAAATTTAATAAAGATGAAAGAACGCTTTTGAATAAATAATTAATAAGAC 2015
QY 1801 gaaaataaacccagatgaaaaaattctagaagatctgacatctgtaaaatgaaatgaa 1860
Db 2016 GAAAAATTAACCAAGATGAAAAAATTTAGAAATATCTGACATTTGTAAGTACAAAGTTCA 2075
QY 1861 aaagtcttaataatgaaacaaatctgaagaatttaaaaaagactcaattgatttaaaaaa 1920
Db 2076 AAAGTTTATTAATGAAACAAATTTGACGAATTAATAAAGACTCAATTTGATTTAAAAAAT 2135
QY 1921 gtaagaattaaacataataataatgattcccaatctctacaaagaagaataaagaa 1980
Db 2136 GTAGAAATTAACATATATATATGTTCCCAATTTCTTACAAACAAATAAATAGCAAGAA 2195
QY 1981 ccttattacttaattgcttgaaaaaagaattgataaattaaagtttcatgacctaaag 2040
Db 2196 CCTTATTTATTTATTTGTTGTAAGAAAAAGAAATTTGATTAATTAATAATGTTCAATGCTTAAG 2255
QY 2041 gtaagaatctgataaataatgaaagaaaaaataaataaagaagaagatcgaatgaaat 2100
Db 2256 GTAGAAATCTGTTGTAATGAAAGAAAAAATAATTAATAAACAACAGCTCAATCGATTAAT 2315
QY 2101 tgggaacccataacccgaaggaataataacaggaacgaacttaaaacctggaacaa 2160
Db 2316 TCGGAACCATCAACCGAAGAAATTAACAGGCAAGCAACTTACAAACCTTGACAAACA 2375
QY 2161 gcaagatctgctttagaagagatctcaatcaagcaagaacacaaagaaacaaacaa 2220
Db 2376 GCAGAGATCTGCTTTAGAGAGATTCATGATCAAGCAACAAACAAAGAAACAAAGCA 2435
QY 2221 caaccacagttaccagttaccagttaccagttaccagttaccagttaccagttaccag 2280
Db 2436 CAACCAACCATACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 2495
QY 2281 ccagtaataataaactgaaatgcttccaaattagatlaattgaaaaatataatga 2340
Db 2496 CCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2555
QY 2341 ttttaaatctcatatataatgctcaaaatatatttggtttcaaacatcaatgaac 2400
Db 2556 TTTTAAATCTTCAATATATATGTCACAAATTAATTTTGGTTTCACACTCAACTATGAAC 2615
QY 2401 gaaagataatlaaaacataataaattacaaaggaagaaagaaacaaatlaaattcaat 2460
Db 2616 GAAAGATTAATAACAAATTAATAATTTACAAAGAGAGAAAGAAATTAATTAATTAATTA 2675
QY 2461 gatccatgaacttatttggttataatacaaaataacataactgtaattctatgctt 2520
Db 2676 GATCCATTAAGCTTATTTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2735
QY 2521 gatagcttaaaacatgatttatacaactat----- 2551
Db 2736 GATAGCTTAATAC-ATAGCTTATACAACTATTTATGCGTTTATGAATAAAGAAATTTGTTA 2794
QY 2552 -----tlatgaaaaattatgaaaaagaatggttctgtaattatataaact 2597
Db 2795 TGATTTATATATCTTATATGAAATTTATGAAAAAAGAAATGTTGTAATTTATATTAAC 2854
QY 2598 taagataatgacaaataataaaattatataagaagaagcaaaagatccaactctg 2657
Db 2855 TAAAGATTAATGACAAATAATTAATAATTTATAGAGAGAGAAAAAGTATCCACATCTGT 2914
QY 2658 aaaaactcttcaagttcatcaatgcaacatatacatatacaactcaagaaataaccca 2717
Db 2915 AAAAATCTTTCAAGTTCATCAATGCAACCATTTATATTAACCTCCAGATTAACCCCA 2974
QY 2718 agtaagttgcaaatgatacatatcacaattcttaaaattgaaatlaattgaaatla 2777
Db 2975 ACTAATGCAAAATGATGATATACATCATCTTCACAAATTTGAATTAATTAATTAATTA 3034
QY 2778 tgaatacatatagagcttggaaaaaaataataataccaaagatlaattggttcaaa 2837
Db 3035 TGAATAATATTTGAGCTTGGAAAAAACAATAATATATACCAAGATTAATTAATTAATTA 3094
QY 2838 aagtaatgaaacattatgaaagaataataaagaatgataatcttataaagaatc 2897
Db 3095 AATGATGAAACCTTTATGAAAGAAAGATTAATAAAGATAGTATGATTTTATATATTAAT 3154
QY 2898 tttaacaaatttgtaaaatctaaagctgataatlaattcaattgaatgaaatcaaa 2957
Db 3155 TTTTACAAATTTGTAATAATTTAAAGCTGATGATTAATTTCAATGAATGAATCAAA 3214
QY 2958 aaggaagaattagaagaatataataaattaaaaaaactctacagttatcaattga 3017
Db 3215 AAGGAAGAAATTAAGAAAGATTAATTAATTAATAAATAAATAAATAAATAAATAAATAA 3274
QY 3018 ttataataataataataataatlagaaagatatttgtaaaagaagaagaagttg 3077
Db 3275 TTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3334
QY 3078 taataataatgcaaatlaaaataacttactatlaaaagaacaaatlaagaatccaact 3137
Db 3335 TAAATTAATAAAGCAAAATTAATAAATACTTTATTAATAAAGAAATTTGATTAATTAAT 3394
QY 3138 gaattcaacttaataacccaagacatgatatcaaaaacttctggttcttcaacaaa 3197
Db 3395 GAATTCACCTTAATTAACCAACAGATGATTAACAAACCTTCTGTTTCTTTAACAATAA 3454
```


OY	31198	aaaagaagctgnaa	tagcaaaaactgnaaacacattaga	aaacacaaatactatctga	3257			
Db	34455	AAAAGAAGCTGA	AAATPACCAAAATCAGAAAAACATTG	AGAAAAACAAAATATTATTGTAA	35144			
OY	3258	acattataaaggact	ctctaaatcata	tggtagcatctccattaaaccttaag	3317			
Db	3515	ACATTATTAAGAC	CTTGTATAATTATATA	TGGTGAATCATCTCCATTAAAACTTTAAG	3574			
OY	3318	tgaagaatcaat	lccaacagaagaata	atlatgcca	gttagaanaacttgaaglatia	3377		
Db	3575	TGAAGAATCAAT	TCAACACAGAAATATTATG	CGATTATGAAAACTTTAAGGTATTAAG	3634			
OY	3378	taaatagaaggnaa	atltaaagata	atlltaaat	ttagaanaaagaaaatactact	3437		
Db	3635	TAAATTTGGAAG	GAATAATTAAGATATCTAA	ATTTTGAAGAAAAAAATATTATCTACTT	3694			
OY	3438	atcaagtggaat	atacatctaa	ctctgaat	taaaagaaagatataaaaataaata	3497		
Db	3655	ATCAAGAGGTTT	ACATCATTTTAA	TGCTGAATTTAAAGAAAGTA	TAAAAATAAAAATTA	3754		
OY	3498	tacaggttaat	ctctccaagtg	aaaaataca	cgatgtltaacaa	tgatltgaactlta	3557	
Db	3755	TACAGGTAAAT	CTCCAGGCTAAATATAG	CGAGTGTAAACATG	CGATTAGAAATCTTTACAA	3814		
OY	3558	aaaattctcccaga	aagacagag	ctgccaag	ctgtgaag	tgaaagtgcgcgac	3617	
Db	3815	AAAATTTCTCC	AGAAAGAACAGATGTTGC	ACAGTTGTAGTGAAGAGCGATCCGACAC	3874			
OY	3618	atagaacaaagt	caacaaagaaac	cgatcaact	ctatgtaga	gcagagcttcaaac	3677	
Db	3875	ATTGGAACAAAG	CTCAACCAAGAACCGATCA	CTCACTGTATGAGACGCA	GACTCTCAACAC	3934		
OY	3678	aatacacat	caacaaga	ctgcgatga	agtagaga	tgcgtlaa	ccaatga	3737
Db	3935	AATTAACAACAT	CACAAAATCTCGATG	TACAGTAGATGACGTA	ATCATATGATCTCATATT	3994		
OY	3738	tgaggaa	lccgaagaaga	atga	tgat	taggacaag	tagtaacaggaagacga	3797
Db	3995	TGGAGAATCCGA	AGAAAGATATTATGATGATT	TAGACACAGTAGTAAC	GAGGACAGTAGTAC	4054		
OY	3798	tccttcgcg	taattgata	actctcttaaat	ctgaaatg	atgatatgag	gtttatatt	3857
Db	4055	TACTTCGGTAT	TGATTAACACTACTTCT	TAAAAATGAAATGATAT	TGAGGTTTATATT	4114		
OY	3858	aaaacctga	gcag	gtgttata	gaaggtt	taaaaaaacaat	tagaabaat	3917
Db	4115	AAAACTTTAG	CAGGCTGTTTATATG	AAAGTTTAAAAAACAT	TAGAAATACGTTATAC	4174		
OY	3918	atttaatgt	taatgtt	aagata	tttlaaat	lccagat	ttlaa	3977
Db	4175	ATTTAATGT	TAAATGTTAAGGATATT	TTTAAATTCAGATTTAA	TAAACGTAAAAATTTCA	4234		
OY	3978	aaatg	tttaga	atccga	tttaatt	cccatata	aaga	4037
Db	4235	AAATGTTT	TGGAATCA	CAATTTAATTCATATA	AGATTAAAGTTTAAACAT	TACAGTAATTA	GTGT	4294
OY	4038	caaaagct	ccata	taaa	ctcttata	aagaaaaaag	aagataa	4097
Db	4295	CAAAAGATCC	ATATAATTTTATATA	AGAAAAAGATTAATTTCT	TAAACGATTATA	4354		
OY	4098	ttatcat	aagga	ttca	tagata	cgata	laaatt	4157
Db	4355	TTTATATTAA	GAGATTCAATACAT	CGGATPAAATTTTG	CAATTAATGATGTTCTT	GGATATTTA	4414	
OY	4158	taaaa	attat	ccgaaaaa	ataata	ccagat	ttagat	4217
Db	4415	TAAATATT	TTCGAAAAAATTA	ATAACAGATTTAGATTCT	TAAAAAAATATATAC	MG	4474	
OY	4218	caaa	caggtg	aaaa	tg	aagaa	ta	4277
Db	4475	CAAAACAAG	GTGAATAATG	CAAAATACCTTCC	CTTTTAAACAATAT	TGAGACCTTATATA	4534	
OY	4278	aacag	ctta	atgata	naaat	tg	atattat	4337

DB	LOCUS	DEFINITION	ACCESION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	REFERENCE	AUTHORS	TITLE	JOURNAL
Db	4535	AAACGTTATGATTAATAATTGATTTATTTGGTAATTCATTTTAAAGCAAAAGTTCTTAATAA	4594													
QY	4338	taacataatagaatacaaacgtagaagcttaaaataaagaacttaacttaaaacaat	4397													
Db	4595	TTCATATGGAAGTAATCAACGTTGAAGTTTAAATAAAGAACTTAATTCTTAAAAACAT	4654													
QY	4398	taagaacaaatlygcagatlttaaaaaataacaattcgtltygaattgctgatlac	4457													
Db	4655	TCAAAGCAAAATGGCAGATTTTAAAAAAATPACAAATTCGTTGAATTCGATTTATC	4714													
QY	4456	aacgaattataacccataactatltgacaaagttccttltgtaagtaatggttttga	4517													
Db	4715	AACCAATTAATAACCAATTAATTAACCTTTACAAAGTTCCTTGATACAGGTATCGTTTTGA	4774													
QY	4518	aaattctgtaaacccgttttcttaacttactga-ttggaaccttgcaagtaatgta	4576													
Db	4775	AAACTTGCTAAATCCGTTTATCTTAATTTACTTGTATTTGGAACCTTGCAAGGTATGTTAA	4834													
QY	4577	acatttcaacaacccaatgcygtlaaaaaacaatg--tcacaaaatltctgagttltag	4634													
Db	4835	ACATTTCAACACCAATGCGTAAAAAACAATGATCCACAGAGTTCTGGATGTTTCAG	4894													
QY	4635	acatttagatgaaagaagaatgtaaatgltatlaaatacaaaagaagaatgataa	4694													
Db	4895	ACATTTAGATGAAGAAAGAGAGATGTAAATGTTTATTAATTAACAACAAAGAGATGATA	4954													
QY	4695	atgitttgaaaatccaaatcttacttcttgaagaataatgltgattgtlgtcagatgc	4754													
Db	4955	AGTGTGTGAAATTTAAATCTCTACTTGTACCAAAATATATGATGTGATGTGACGATGC	5014													
QY	4755	caaatgtacccgaagaagatltcagtagaacaagaaagaanaatacaatgttgatgataa	4814													
Db	5015	CAATGTACCGAAGAGATTCAGGTACGACACGGAAGAAATACATGTGATGTACTTAA	5074													
QY	4815	accgtatcttataccacttttcgatgtgataatlttcgcagttcctctaacttcttagaat	4874													
Db	5075	ACCTGATTTGTTATTCACATTTTCGATGTGATTAATTTTCGCGATTCCTTACTTCTTGGAAT	5134													
QY	4875	atcattcttataataaccatggtataatataatacagtttcattaa	4920													
Db	5135	ATCATTTCTTATTAAATCATCATGTTATATTATTATACAGTTTCATTTTAA	5180													
RESULT	6															
PEP190		5282 bp	DNA	INV	21-FEB-1996											
LOCUS	PEP190															
DEFINITION	P.falciparum gp190 (MSA1, MSP1, PMMSA) gene for merozoite surface															
ACCESSION	X03371															
VERSION	X03371.1	GI:929797														
KEYWORDS	gp190 gene; merozoite surface antigens; MSA1 gene; MSP1 gene; PMMSA															
SOURCE	gene.															
ORGANISM	malaria parasite P. falciparum.															
REFERENCE	Plasmodium falciparum															
AUTHORS	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.															

der, Universitaet (ZMBH), Im Neuenheimer Feld 282, Heidelberg
69120, FRG
On Aug 5, 1995 this sequence version replaced gi:9920.

COMMENT Location/Qualifiers

FEATURES

source

1..5282

/organism="Plasmodium falciparum"

/isolate="K1"

/db_xref="taxon:5833"

/chromosome="9"

/dev_stage="blood stage"

72..128

/gene="gp190, MSAL, MSP1, PMMSA"

/product="putative"

72..4964

/gene="gp190, MSAL, MSP1, PMMSA"

72..4964

/gene="gp190, MSAL, MSP1, PMMSA"

/codon_start=1

/product="precursor for the major merozoite surface
antigens"

/protein_id="CAA27070.1"

/db_xref="GI:929798"

/db_xref="SWISS-PROT:P04932"

/translation="MKITFLCSFLFEIINTOCVTHESYQELVKLEALEDAVLGYS
LEHKEKMLNEBEITTKGASQSGTSGSPSSRSNTLPBSNTSGA
SPPADASDASADLKHRYNYLTITELKYPOLDLTNHMLTCDNIHGFYLD
GYEINELLYKLNFEYDLRAKLDVCAVCOIPENKIRANEDLVKLVGRKP
LDINQDNQKMDYIKKKNTIENELNEESKKTINKKPPRANSNTPTLKD
STYKQLEPAHNLISVLEKRIIDTLKKNELKIDKINKKPPRANSNTPTLKD
NKKIEHEKEIKELATIKFNIDSLDLELYLEKYNKKNIDISAKETESTEPN
EYKPVNTYPLSYNDINMLNELNSFGDLINPDYKPEKKNITDNEKFTNEIKK
IKIEKKIESDKSYEDRSKSLNDITKEKELNLYEYKNNIDLTNEKMGKRY
SYKVEKTHNTFASYSNHNKELTKAKYEDSLNLYVEKLYKKNLSKIE
NEIETLVENIKKDEOLPEKRIKDEKNEDEKILEVSDIVKVOVOKVILMKIDELK
TOLLINKVELKHNHVPNSYKQNEKPEYLLIVKKEIDKLVPAKPVESLNEEKN
ITTEGSDNSPSTBEITGQATTKPGQAGSALBDSVOAOAODKQAPVPAVP
EKAQVPTPAPVNNKTEVNSKLDVLEKYEFLNTSYICHKYLIVSHSTNKKILKQY
KITKEESKLSGCDPLDLFNIONNIPWYSMEISLNNLSYLFMEIYERKEVNCILKQY
LKDNDRKMLLEAKKVSSTVKLSSSMQPLSLPODPEVSANDTSHNNLNSL
KLEENILSGKNKNIYQELIGOKSSBNFEKILKSDPPEVSANDTSHNNLNSL
NDESKRKLEEDINIKLTIQSLDPLYNKIKLIRPEKKTGVKQKQIKKTLTK
EOLSKLNSLNPKVILONFVFNKKKAELAEENTLENTILKIKHKKLVKTYNG
ESSPLKTSUESIQTEDNYSLENFVKLSLEGLKLDNLEKKIYSTLSGLHILIA
ELKEVITKNKNSPSENNITNNALEBKKEFLPEGDVATVVSSESDTLEOSQK
KPSHNGEASNTITTSQVNDVEDDVIIVPFGESEEDYDGLGVTEAVTPSYD
NLSKIEHNEEVLVYLKPLAGVRSKLKOLENNWFMENVVVKDILMSREKREMFNVL
ESDLPYKDTSSNVVDPKFKLEKRDKLSYNIKSDIDPIDINANVLYGK
ILSEKYSDDLSTIKRYINDQGENEYLPFLNNIEFLYKTVNDKIDLEVIHLEAVLN
YTERKSNVEVKIKELNYLKTIDOKLADFERKNNFVIGIADLSDYHNNMLTFLSTGM
VEENLAKTVLSNLDGNLQGMINIISOHQCVKQCONSGCFPHLDRECECKLLNKO
EDCKVENNPCTNENNGCGDADAKCTEEDSGSNGKKITCECTKPDSPYLPFGICSS
SMFISIFLLIMLILYSFI"

BASE COUNT 2363 a 608 c 676 g 1635 t
ORIGIN

Query Match 96.6% Score 4750.4; DB 96; Length 5282;
Best Local Similarity 98.1% Pred No. 0;

Matches 4827; Conservative 0; Mismatches 66; Indels 27; Gaps 1;

QY 1 atgaagatacatctcttattgltcatctcttcttattataataacacaaatgta 60
DB 72 ATGAAGATCATATCTTTTATGTCATTTCTTTTATATTAATAACACATGTGTA 131
QY 61 acacatgaagaatatacaaaactgtcaaaaactagaagctttagaagatgagatg 120
DB 132 ACACATGAAGAATATCAAAACTGTCAAAAACCTAGAAGCTTTAGAAGATGAGATTTG 191
QY 121 aatgaatcagaatattatcaaaagaaaataatgataataaagaagaacagtgaaaca 180
DB 192 AATGAATCAGAAATGATAAGGAAAAATGATCTTAATGAAGAAGAAA----- 243
QY 181 gctgttacaactaglacacctggttcaaggtctcaggtcaggtggtcaggtgagc 240

DB 244 -----TTACTACAAAAGGTGCAAGTGCTCAAGTGCTACAAAGTGT 284
QY 241 tcaagttcagcaggtggtcaggtggtcaggtggtcaggtggtcaggtggtcaggt 300
DB 285 ACAAGTGTGTAACAAGTGTGTCACAGTGTGTCACAGTGTGTCACAGTGTGTCACAG 344
QY 301 gcttcaggtggtcaggtggtcaggtggtcaggtggtcaggtggtcaggtggtcag 360
DB 345 ACTTTCAGTGTGTCACAGTGTGTCACAGTGTGTCACAGTGTGTCACAGTGTGTCACAG 404
QY 361 gatgctaaatcctcagcgtgatttaaaacacagagtaacgaatctctgttaacatcaaa 420
DB 405 GATGCTAAATCTTACGCTGATTTAAACACAGAGTGTGTCACAGTGTGTCACAGTGTGTCACAG 464
QY 421 gaactcaataatcccaactcaggttgaatcaatcaggtggtcaggtggtcaggtggtcag 480
DB 465 GAACCTAAATVTCCTCAACCTGATTTAACTAAATCAATVTCCTGATTTAACT 524
QY 481 attcaggttcaaatatttaattgattgattgattgattgattgattgattgattgattgatt 540
DB 525 ATTCATGCTTTCAAAATTTTAAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 584
QY 541 ttaactttatttgaattatataagaacaaataaataatgaatgattgattgattgattgatt 600
DB 585 TTAACCTTTATTTGATTTATTAAGAGCAAAATTAATGATGATGATGATGATGATGATGATGAT 644
QY 601 tgcataatccttcaactcaactcaactcaactcaactcaactcaactcaactcaactcaact 660
DB 645 TGCAATATACCTTTCAATCTTAAATTCGCAAAATGAATTAACCTTACTTAAATTAACCT 704
QY 661 gttctggatataagaacacatcagcaaatataagaataatgaagaataatgaagaata 720
DB 705 GTTCTGGATATATGAATAACCATTAACCATTAATTAAGATTAATTAAGATTAATTAAGATTA 764
QY 721 tacatttaaaaaataaaaaaacatagaataataatgaatgaatgaatgaatgaatgaat 780
DB 765 TACATTAAAAAATAATTAATAAACCATTAAGAAATTAATTAATTAATTAATTAATTAATTAAG 824
QY 781 aaaaacattgataaaataagaatgtaacataagaagaagaagaagaagaagaagaagaaga 840
DB 825 AAAACATTGATTAATAATTAATAAGTGAACATAAGAAAGAAAGAAAGAAAGAAAGAAAGTATACCAA 884
QY 841 gctcaatgatctctctatcttcaataaacaattagaagaacacataatcaataagc 900
DB 885 GCTCATATGATCTTTCTATTTTCAATTAACATTAAGAACATTAATTAATTAATTAATTAAGC 944
QY 901 gtttgaagaacgattgacacattcaaaaaaaatgaagaacataaaggaaatcttga 960
DB 945 GTTTTGAAGAAACGATVTCACACTTTAAAAAATAATGAAGAAACATTAAGGAATTAATGAT 1004
QY 961 aagataataatgaataataaaatccccccacggccaattctggaataatcaccaaatcttc 1020
DB 1005 AAGATTAATGAATTAATAATTAATAATCCCGCCCGCAATTCGGAATTAACCAATTAATCTTC 1064
QY 1021 ctgataagaacaaaaataatcagagagacacgaagaagaataaagaagaataatgccaact 1080
DB 1065 CTGATTAAGACAAAAAATTCAGAGACACGAAAAAGAAATTAATAAATAATTCGCCAAAACCT 1124
QY 1081 attaaattcaatattgattgatttatttatttatttatttatttatttatttatttatttatt 1140
DB 1125 ATTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1184
QY 1141 gaaaaataataaataatgataaagtgcaaaaggttgaacaaagaagaatcaactgaacc 1200
DB 1185 GAAAAAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1244
QY 1201 aatgaatcagaatcagaatcagaatcagaatcagaatcagaatcagaatcagaatcaga 1260
DB 1245 AATGAATATCCAAAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1304
QY 1261 aatgaatcaatctcttctggtatttaataatcaattgattatatacaaaagaacaggt 1320

Db	1305	AATGACTTAATTCCTTTGGTGATTTTAATTAATCCATTTTGATTTATACAAAGACCAAGT	1304
Oy	1321	aaaaacataatactcgtgataa tgaagaagaaaaaaat tcaataa tlgaaat tgaagaaaaa at	1380
Db	1365	AAAAAATATATACTGATTAATGAAAGAAAAAATTCATTAATGAAATTTAGGAAAAAATT	1424
Oy	1381	aaat tgaagaaaaaaaat tlgaaat cgtataa aaat cttacgaagacagatctaagctt	1440
Db	1425	AAAAATAGAAAAAATAAATTTGAAATCTGATTAATAAATTTACGAAGACAGATCTAAAGCT	1484
Oy	1441	ctaaatgtatcaacaaagaata tlgaaaaaat ttaactaaat tgaattat gatata gaaatc	1500
Db	1485	TTTAATGATTAATACAAAAAGATATGAAAAATTTACTTAAGAAATTTATCATAGCAAAATTC	1544
Oy	1501	aataataatagatcttcaactaaat ttcgaaaaaat gat tggtaaaagat tcatataa	1560
Db	1545	AATATATATATAGTTTAATCTAATTTTCGAAAAAATGATGGTAAAGATTTTCATATATAA	1604
Oy	1561	gt tgaagaacttaacacacataa tactttgtcaatccat tlgaaat tcttaacaataatctt	1620
Db	1605	GTTGAGAAACTTACACACCATATATACCTTTTGCACTGATGAATAATTTAAACATTAATCTT	1664
Oy	1621	gaaaagttaacaaaagctctctaataat tggaaagat ttaactctt ttaagaata tagtagt	1680
Db	1665	GAAAAGTTAACAAAAAGCTCTTAATATATAGGAAGATTAATCTTTAAGGATATATAGTACTT	1724
Oy	1681	gaaaagaatctaaaat tcttaaaaaat ttaaaagaagaaatagaanaat tgaagt tgaaca	1740
Db	1725	GAAAAAGATTTAAATATTATTAAAAATTTAATTAAGCAAAATAGCAAAATGAGATTGAACA	1784
Oy	1741	ttag ttgaanaat ttaaaaaaga tgaagaacagctt tctgaaaaaaaat ttaactaaagac	1800
Db	1785	TTAGTGAATATATTAAAAAAGATGAAGAACACTTTTGAAAAAATAATTACTAAACAC	1844
Oy	1801	gaaaaataacccagat tgaaaaaat ttttaagaat ctcgacat tgtataa agtataa agtctaa	1860
Db	1845	GAAATTAACACAGTGTAAAAAATTTTAAGAAAGTATCTGACATTTGTAAAGTACAAATGTCAA	1904
Oy	1861	aaagt ttaataa tgaacaaaat tgcgcgaat taaagaagcttcaat ttaataaat	1920
Db	1905	AAAGTTTATTAATGACCAAAATTTGACGATTTAAAAAACAATTCATTTTAAAAAAT	1964
Oy	1921	gtagaat taaacataa tataa taca tgtcccaat tcttaacaa caaagaanaat tgaagaagaa	1980
Db	1965	GTAGAATTTAAACACTATATATACATGTGCCAATTTCTTAAACAAAGAAATTAAGCAACAA	2024
Oy	1981	ccattatcttaat tgttgt tgaaaaaagaat tgaataa ttaaaagt tctcaagctcctaag	2040
Db	2025	CCCTTATTTATTAATTTGCTTGAAAAAAGAAATTTGAATTAATTTAAAGTGTCATGCCCTAAG	2084
Oy	2041	gtagaatcat tgtttaa tgaagaaaaaaaacataa aaacgaagaagtcacatcgatataat	2100
Db	2085	GTAACATCATATTGATTAATGTGAGAAAAAATAAACATTAATAAACGAAAGTGCAATCGGATATAT	2144
Oy	2101	tcgaaaccat tcaacccggaaggaagaat taaacagga caagcaact tacaanaa cct tgaacaacaa	2160
Db	2145	TCGGAACCATCAACCCGGAAGAGAAATTAACGAGCAACGACAACTACAAAACCTTGACAAACAA	2204
Oy	2161	gcagagatctgcct tgaagaggat tcaagtaacaa gcaacgaacaa gnaacaaaacaa gca	2220
Db	2205	GCAGAGATCTGCTTTAGAGAGAGATTTCAATCAAAAGCAACAGCAACAAAGAACAAAAACAACCA	2264
Oy	2221	caaccaccag taccagracacag taccgaagaagcaaaagccaaagt tcccaaccaccacagca	2280
Db	2265	CAACCCACCAAGTACCAAGTACCAAGCAAGCAAGCAAGCAAGTCCCAACACACACACACCA	2324
Oy	2281	ccagt taaataa taaact tgaanaa tgttccaaat tgaat tcatct tgaanaa ttaata tga	2340
Db	2325	CCAGTAAATATATAAACTGAAAAATGTTTCCAAATTTGATTAATCTTGAAAAAATTAATATGAA	2384
Oy	2341	ttttaataatct tcatataa tcat tgcacaaatataat tttgttttcaacactcaacta tgaac	2400
Db	2385	TTTTTAAATATCTCATATATATATGTCCAAATATATATTTTGTTTCACAGCAACATGTGAC	2444

OY	2401	gaaaagatatlaaaacaatatlaaaatlcagaagaggaagaaagcaaatlaagtlcatgct	2460
Db	2445	GAAGAAGATATTAAACAAATATTAATAATTACAAAGAGGAAGAAAGAAATTAAGCTTCAGCT	2504
OY	2461	gacccatlaagcctatcgtttlaatatlacaaaataaacacaccgtbaatlaatacctcgttt	2520
Db	2505	GATCCATTAGACTTATGTTTAAATATACAAAAATACATACCTGTATGTATTCTATGTTTT	2564
OY	2521	gatagcttaaacaaatgattlacacaaacttbtatygaaattlatgaaaaagaatlgct	2580
Db	2565	GATAGCTTAAACATAGCTTTATATCACCAACTTTATTGAAATTTTGAAGAAACAAATGSGTT	2624
OY	2581	tgtaatttatataaactlaaggaatagcacaatatlaaaatttctttagaggaagcgaa	2640
Db	2625	TGTAAATTTATATAACCTTAAGATTAATGACAAAATTTAAAAATTTATTTAGAGGAACCGGAA	2684
OY	2641	aaagtlaccacatctgtaaaaaactcttcaagtlctcaatgcaacatlatcatlaaca	2700
Db	2685	AAAGATCCACATCTGTAAAAACTCTTTCAGTTCATCAATGCACACNTATATCATTTAACA	2744
OY	2701	ccccaagataaaacccgaaglaagtgcaaatgatalacatlcacatcttacaattlgat	2760
Db	2745	CCTCAGCATTAACCCCGAAGTACGTCAAAATGATGATACATCATCTTCACAAATTTGAT	2804
OY	2761	aatgctttaaattatcttggaaaaacatltgaagctcttggaaaaaacaaaatatatacaa	2820
Db	2805	AATGCTTTAAATATTATTGGAACATATTTGAGTCTTTGGAAAAACAAAATATATACCA	2864
OY	2821	gaatlaatlaagtlcaaaaaagtagtgaanaacttlatlgaaaaagatalctaaagatagtgat	2880
Db	2865	GAATTAATAGCTCAAAAACAGTAGGAACCTTTATGAAAACATATTTAAACATGTAGTAT	2924
OY	2881	acatttcatatgaatcttltcacaaattltgtaaaaatctcaagctgagatataatcca	2940
Db	2925	ACATTTTATAATGCAATCTTTTACAAAATTTTGTAAAACTGACATATTTAATTTCA	2984
OY	2941	ctgaatgtgaatccaanaagaaagaattagaagaagatlatlaataatlaaaaaaat	3000
Db	2985	TTGATATGATGATCAAAAAGAGCAAAATTGAAAGAGATTTATTAATTTAAAAAAACT	3044
OY	3001	ltacagttatbcatltgatttataataataataataataataatagaagaatlatlgat	3060
Db	3045	TTTACGTTATCATTTGATTTATTAATTAATTAATTAATTAATTTGAAAGATTTATTTGAT	3104
OY	3061	aaaaagaanaagctltgtaaataataaibgcaattlaaaaaacttacttataaagaa	3120
Db	3105	AAAAAGAAACAGTTGTATTAATATTAANAATGCAAAATTTAAAAACTTACTTTATTTAAAGAA	3164
OY	3121	caatltgaatlcanaaatltgaaatlcaactlaataaccaagacatglatlcaaaaaacttct	3180
Db	3165	CAATTAGATCAAAAATGGAATTTCACTTAATTAACCCAAACATGTTATPCAAAACCTTTCT	3224
OY	3181	gtttcttcttaaaaaaaaagaagcttgaatlaagcagaacitgaaaaacacatltgaanaac	3240
Db	3225	GTTTCTTTTAAACAAAAAAGAGAGCTGAAATATGCGAAATCGAAACCATTTACAAAAC	3284
OY	3241	acaaaaatattatltgaaacaatlaaagaagcttgttaaatatlatlaatgltgaaatcatct	3300
Db	3285	ACAAAAATATTATTGAACACATTAATAAGAGCTTGTTAATATTATTAATGTGAATCATCT	3344
OY	3301	ccatlaaaaaactltlaagltgaagaatccaatltcaaacagaagaataatlatgcccgtttgaag	3360
Db	3345	CCATTTAAAAACCTTTAACTGAGAGATCAATTTCAACAGCAAGATTAATTTATGCCGTTTACAA	3404
OY	3361	aacttlaaagcatlaagtlaaatagaagaagaaatlaaagatlaattilaattlaaattlaagaag	3420
Db	3405	AACTTTAAAGATTAACTTAATTTGAAAGCAAAATTTAAAGATTAATTTTAAATTTTAAAGAAAG	3464
OY	3421	aaaaaatatcatcatlatacaagtggaatlaacatlcattlaattgtcgaatlaaagaagla	3480
Db	3465	AAAAAATATCATCTTAATCAAGTGAATTAACATCATTTTAAATTTGCAATTTAAAGAAATGA	3524

OY 3481 ataaaaataaataatatacagglaattctccaagtgaataataacga tgttaacaat 3540
 |||
 Db 3525 ATAAAAATTAATTAACAGTAATTTCTCAAGTGAAATATATACGATGTTAAACAT 3584
 |||
 OY 3541 gcatltaaatcttaacaaaaattctccagaaggaacagatgttgcacagtgttaagt 3600
 |||
 Db 3585 GCATTAGAACTTACAAAATTTCTCCAGGAAGAAACATATGTGCAACAGTGTGAAGT 3644
 |||
 OY 3601 gaagaagtcacacacattagaacaaagtcaaccaaagaacacagataactcatgta 3660
 |||
 Db 3645 GAAAGTCGATCCACACATATTAGAACAAAGTCAACCAAGAACACGATCACTCATGTA 3704
 |||
 OY 3661 ggaagcagatctcaacaatacaacatacaaaatgtcgaatgaatgaatgaatgaatga 3720
 |||
 Db 3705 GGAGCAGAGCTTAACACAAATTAACACATCAACAAATGTGATGATGATGATGATGATGATG 3764
 |||
 OY 3721 atcatagtaactatatttgaagaatccgaagaagatataatgaatgaatgaatgaatga 3780
 |||
 Db 3765 ATCATAGTACCTATATTTGGAGAAATCCGAAGAAATATATATATATATATATATATATATAT 3824
 |||
 OY 3781 acagaggaagcagtaactcctccgaatgaatgaatgaatgaatgaatgaatgaatgaatga 3840
 |||
 Db 3825 ACAGGAGAAACATACTCTTCCGTAATGATACATCTTTCTTAATTAATTAATTAATTAATTAAT 3884
 |||
 OY 3841 tatgaagtttataatataaaccctttagcagtgattatagaagtttaaaaaaacaata 3900
 |||
 Db 3885 TATGAGGTTTATATATTAATAAACCCTTAGCAGGTGTTATGAAAGTTTAAAAAAACATTA 3944
 |||
 OY 3901 gaataaagctatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 3960
 |||
 Db 3945 GAAATAACCTTATGACATTTAATGTTAAGTTAAGATATTTTAAATTCACCATTTAAT 4004
 |||
 OY 3961 aaacgtgaataatctcaaaaatgtttagaagaatgaatgaatgaatgaatgaatgaatga 4020
 |||
 Db 4005 AAACGTAATTAATTCAAAATGTTTGAATGATTAATTAATTAATTAATTAATTAATTAATTAAT 4064
 |||
 OY 4021 tcaagtaaatatgttgcagaatgccatataaattcttaataaagaagaagaagaagaaga 4080
 |||
 Db 4065 TCAAGTAATATGTGTGCAAAAGATCCATTAATTTCTTAATTAAGAAAGAAAGATTAAT 4124
 |||
 OY 4081 tcttaagcagttataatataatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 4140
 |||
 Db 4125 TTCTTAAGCAGTATATATATATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 4184
 |||
 OY 4141 gatgtcttggaatataatataatataatataatataatataatataatataatataatata 4200
 |||
 Db 4185 GATGTCTTGGAAT 4244
 |||
 OY 4201 aaaaaataatacaacgaacaaagtgaaatgaagaatgaatgaatgaatgaatgaatgaatga 4260
 |||
 Db 4245 AAAAATATATCAACGACAAACAAAGTGAAATGAGAAATACCTTCCCTTTTAAACAT 4304
 |||
 OY 4261 attgaagcattataaacaacgttaaatgaatgaatgaatgaatgaatgaatgaatgaatga 4320
 |||
 Db 4305 ATTGAAACCTTAT 4364
 |||
 OY 4321 gcaaaagcttaaatataatataatataatataatataatataatataatataatataat 4380
 |||
 Db 4365 GCAAAAGTCTTAAT 4424
 |||
 OY 4381 aattacttaaaaaacaatcaagaacaattggaacatttaaaaaaatacaaatctcgtt 4440
 |||
 Db 4425 AATTACTTTAAAAACATTCACGACAAATGGCGATTTTAAAAAATTAACATTTCCGTT 4484
 |||
 OY 4441 ggaattgctgattatcaacagaatataacataaataactatttgaacaaagtcctagt 4500
 |||
 Db 4485 GGAAATGCTGATTTTACACAGATTTAATCAATTAATTAATTAATTAATTAATTAATTAATTAAT 4544
 |||
 OY 4501 aaaggaatgatttgaagaatctgttaaaacggtttatctcaatttacttgaatgaatga 4560
 |||
 Db 4545 ACAGGATGATGTTTGAATAATCTTGTAAACCGTTTATCTTAATTTACTTGTATGGAAC 4604
 |||
 OY 4561 ttgcaagtatgttaacaatttcaacaaccaaagtcgtaaaaaaacaatgtccacaat 4620

Db 4605 TTGCAAGGATGTATTAACATTTCAACAACCAATGCGTAAAAAAACAATGTCCACAAAT 4664
 |||
 OY 4621 tctggaatgttccagacatttagatgaagaagaagaatgaatgaatgaatgaatgaatgaatga 4680
 |||
 Db 4665 TCTGATGTTTCAACATTTTGAATGAAGAAGAAATGTAATGTTTAAATTTACAAA 4724
 |||
 OY 4681 caagaagtgatataatgttctgaagaatcccaatctgtgaacgaataatgaatgaatga 4740
 |||
 Db 4725 CAGGAAGGTATTAATATGTGTTGAATAATCCAAATCTTACTTGTAAACAAATTAATGTGTA 4784
 |||
 OY 4741 tgtgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 4800
 |||
 Db 4785 TGTGATGCAAGATCCCAATATGATACCAAGAAAGATTCAGTGAACGAAAGAAATACACA 4844
 |||
 OY 4801 tgtgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 4860
 |||
 Db 4845 TGTGAATGTAATTAACCTGATTTCTTATCCATTTGATGATGATGATGATGATGATGATG 4904
 |||
 OY 4861 aactctgaagaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 4920
 |||
 Db 4905 AACTTCTTGAAGATATCATTTCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4964
 |||
 RESULT 7
 AF062349 5271 bp DNA INV 09-FEB-2001
 LOCUS Plasmodium falciparum strain HN2 merozoite surface protein 1
 DEFINITION Precursor (msp1) gene, complete cds.
 ACCESSION AF062349
 VERSION AF062349.1 GI:3859844
 KEYWORDS
 SOURCE malaria parasite P. falciparum.
 ORGANISM Plasmodium falciparum
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 5271)
 JOURNAL Jiang, G., Liu, R., Daubenberger, C.A. and Pluschke, G.
 TITLE Sequence analysis of the Msp 1 gene of Plasmodium falciparum from
 Haian, China
 AUTHOR Chung Kuo Chi Sheng Chung Hsueh Yu Chi Sheng Chung Ping Tsa Chin 17
 (5), 294-297 (1999)
 REFERENCE 2 (bases 1 to 5271)
 JOURNAL Jiang, G., Liu, R.Z., Daubenberger, C.A. and Pluschke, G.
 TITLE Direct Submission
 JOURNAL Submitted (30-APR-1998) Molecular Immunology, Swiss Tropical
 Institute, Socinstrasse 57, Basel CH-4002, Switzerland
 FEATURES
 source
 1..5271
 /organism="Plasmodium falciparum"
 /strain="HN2"
 /db_xref="taxon:5833"
 /chromosome="9"
 /country="China"
 /gene="msp1"
 /product="merozoite surface protein 1 precursor"
 <66..>5180
 /gene="msp1"
 66..5180
 /gene="msp1"
 /note="mspi"
 /note="inds spectrum"
 /codon_start=1
 /product="merozoite surface protein 1 precursor"
 /protein_id="AAC72885.1"
 /db_xref="GI:3859845"
 /translation="MKIIFLCSPLEFIINTQCVTHSYQELVKKLELDAVLGYS
 LEQKEMVINTSGTAVTSTPGSGSGSVASGSGSGSVASVAVSGSVASV
 GSGSGSRTPNSDSDSDAKSYADLKRRVNYLTTELKATPELFDLTNNMLTCDN
 IHGRKYLDGTEBELNELLYKLNFTYDLRAKLNDCANSTCYQIPENLITRANEDLVK
 KIVEGYRPLDNIDKONGVKMDYIKRNTTIANINELLEGSRTTIDOKNADNEGRK
 KLYQAVNYPLYNKOLEAHNLSVLEKRIITLANKENIKRLKLIKIDINKPPANS
 NTPNLLDKNKKIEHEKEIKETIATIFENIDSLPTDLELEYLREKNKRVADYTPKS
 DDPKRSVGIIPVYPYNGIVYPLPLDINHSLAANDKNSYCDLNNPDKIKENKIIT


```

1723 gaagaatgagatggaacatagttgaaataataaaagaatgaagaacgcttttggaa 1782
1806 GAAGCTGATATATAAATAATACAGAGAATAAAGAGTAGTAAGCAAAATCTTAGAA 1865
1783 aaaaaaattcctaagaacggaataaaccgagatgaaaaattttagaagatctcgacat 1842
1866 AAAAAA-----TTTAAAGAGCTAACACATTCAGCAATCTCTTCCTTAGAAGATCTGAATTT 1922
1843 gaagaatgacaagttcaaaaagttttaaataagacaataatgacgaatlaaaaaagact 1902
1923 GTAAATATTACAGTACAAAAAGTTTATTATTAATAAAAAATAGAAAGACTTAGAAAGATA 1982
1903 caattgatttlaaaaaatgtagaataaacaataataacatgattcccaattcttaaaa 1962
1983 GAATTTATTTTAAAAAATGCAACACTAAAAAGATGTATTCATGACCAAAATATTATTA 2042
1963 caagaataatgaagaagaaccttatattatgattgattgtaaaaaagaatgtataa 2022
2043 CCACAAAATATAACCAAGAACCATATATTATTAATGTTAAAAAAGAAAGATAGATTA 2102
2023 aaagtgtcattgcttaagtagaatactatgataaatagaataaaaaacata----- 2076
2103 AAGAAATTTATACCAAAAGATAAAGACATGTAAAGAAAGACAAAGCTGCTTTATCAAGT 2162
2077 aaaaagaagtcgaatcggaataatcggaacacataacggaagagaataaacaagaaa 2136
2163 ATTACACAACTTTAGTTAGTGCAGCAAGCAAACTGAACATGGGGGTCACTCCACACAC 2222
2137 ggaactcaaaaacttgacaacaaagcagatctgctttagaaggaatctagttacaaga 2196
2223 ACATTTATCCCAATCAGAGAGAAACGAGATTAACAGAAACAGAGAAAGAAAGAAACA 2282
2197 caagcaca-----agacaacaaacgaacacacacgaagtaacagttacga 2247
2283 GTAGGACACACACAAACAGTAACATTAACATTTCCACACACACACATGACCCCAAAA 2342
2248 gaagcaaaagacacagtcaccaacacacgacgacagtaataataaactgaaatgct 2307
2343 GAAGTAATAAGTTGTGAATAATTCATTAAGATGAACATTAAGATGACCAATTCACAACTTG 2402
2308 tccaatataatctatcttgaataataatatagaattttaaactctatatatagtcac 2367
2403 ACAAAAACAGTTTATCTTAAGAAATTAAGATGAATTTTAACTAATCATTAATATGTCAT 2462
2368 aaatataatttgatttccacactcaactgaacgaataaagaataaaacataaactaa 2427
2463 AATATATTTTATGATCAAACTCTAGATGACCAAAAAATATTATAGGATATTAATCTT 2522
2428 acaaaaggaagaagaacaaatcaagttcagttgattcattagactatgtttaaata 2487
2523 ACTCCAGAAAGAAAAATGATTAATAATCATGATCATTTAGATTTATTTAATATAT 2582
2488 caaataaacaatacctgaatgatttctcattgattgataagctttaaacaataatatac 2547
2583 CAAAAATACCTACTGCTATGATATTCATTATATGATATGATGAACAATGATTTACACAT 2642
2548 ctattataggaatttatagaataaagaatggttgaattatataaacttaagaat 2607
2643 CTCTTTTTCATTTATATCAAAAGGAATGATTATATTATACATATAACATAAAGAGAA 2702
2608 gaaaaaataaataattatataagaagaagcaaaaa-----gtatcacactgttaaa 2661
2703 AATCACATCAAAAAATATTATTAAGAGCAAAAAAATAAATACTGACATCATCTACATCC 2762
2662 actcttcaagttcattcaatgcaacattatcaatacactcagataaaccga---- 2717
2763 AGTCTGGAATATCAACCGTAATPACTGCTCATATCCGAACTCAAGTATTTCCCAAAAC 2822
2718 ----- 2717
2823 CAACAATCAAAATGCATCTTACCATTAACCAAAATGCTTACTGTATCATCTGGTCTT 2882
2718 -----agtaagtcaaatgataatatacactcttaacaaattgaataatgattaa 2772

```


QY 541 ttaactttatttgatttatttaagcaaaatlaaalgatgtgtgtaattat 600
DB 550 ttgaattttctacttcacttgcctaagcccaaaactgaatgacctttgcccattgacctat 609
QY 601 tgcataactccttcaactccttaaatcgtgcaatgaaatagagcgtacttaaaact 660
DB 610 tgcataattccattcaatttgaattgagaccacaaagattgagacctatgaaagatttg 669
QY 661 ggttcttgatatagaaccatttagcaatataatlaaagataatgtaagaaaatggaagat 720
DB 670 gttcttggaatttcgcaagccttcgcacacatgaacacatgctggcaaaagatggaagat 729
QY 721 taacttaaaaaaataaaaaacatagaaaatataatgaatttaattgaagaagtaag 780
DB 730 tttatt 789
QY 781 aaaaactatgataaaaaataagaatgcaactaaagaagaagaaaaaaatataccaa 840
DB 790 aagaccattgacaaaataatgaatgcaaccaaagagaaagaaagaaagattgtaccag 849
QY 841 gctcaatatagtcttctatttcaataaaatagagaagacataatlaataagc 900
DB 850 gcccaattgacacctgtccatttatacaaaacaccttgaaagaccattgacacctgac 909
QY 901 gtttgaagaaacgtatgcaacttaaaaaaaatgaaaacatlaaggaattactgac 960
DB 910 gttactgagaagacgcatagacacccctcaagaaatgaaatgcaaaactgctgac 969
QY 961 aagataatgaaatlaaaaaatccccacgcccactctggaatacaccnaatctc 1020
DB 970 aagattttaaagaaattttaaagacctccgcaagccacttgagaaaccccttaacgctg 1029
QY 1021 ctgtgataaagaaaaaaatcgagagacagaaaaagaataaagaattgccaact 1080
DB 1030 ctggacaagaacaaagaaatagagagacagaaagaaagaaatgcaaaagctgcaaaac 1089
QY 1081 attaaatttaattgatttacttactgacacacacacacacacacacacacacacacacac 1140
DB 1090 atttaatttcaacttgaatttcttcttactgacacacacacacacacacacacacacacac 1149
QY 1141 gaaaaaataaaaaatctgataaagtgcaaggttgaaacaaagaaatcaactgaaccc 1200
DB 1150 ggaagaaatgaagaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 1209
QY 1201 aatgaatcaacaaatggaattacttacttcttctgataaagataatgaactgttca 1260
DB 1210 aatgaattatcccaatggtgacgacgacgacgacgacgacgacgacgacgacgacgac 1269
QY 1261 aatgaacttaattcttctgatttacttacttacttacttacttacttacttacttacttact 1320
DB 1270 aagacactcaattatgcttgcgtgacttgatttgaaccccttgatttgaatgaaagacccct 1329
QY 1321 aaaaacatatactgataatgaaagaaaaaaatcataaattgaatlaagaaaaaatt 1380
DB 1330 aagaaatttcttacacagacatgagaaagaaagatttatacaagaaatcaagaaagatgac 1389
QY 1381 aaaaatagaaaaaaataatgaaatctgataaaaaatcttaagaagaacgacttaagct 1440
DB 1390 aaaaattgaaagaaagaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaa 1449
QY 1441 ttaaatgataaacaagaataatgaaataatcttaataatgaaatgaaatgaaatgaaatgaaat 1500
DB 1450 cttaaaagatattccttaaaagatgaaagaaatgaaatgaaatgaaatgaaatgaaatgaaatgaa 1509
QY 1501 aataaataatagatttaactaatttcgaaaaaaatgaaatgaaatgaaatgaaatgaaatgaaat 1560
DB 1510 aaaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 1569
QY 1561 gttgagaacttaacacataacttctgatacttctgatacttctgatacttctgatacttctgata 1620
DB 1570 gttgagaacttaacacataacttctgatacttctgatacttctgatacttctgatacttctgata 1629

QY 1621 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 1680
DB 1630 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 1689
QY 1681 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 1740
DB 1690 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 1749
QY 1741 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 1800
DB 1750 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 1809
QY 1801 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 1860
DB 1810 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 1869
QY 1861 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 1920
DB 1870 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 1929
QY 1921 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 1980
DB 1930 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 1989
QY 1981 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2040
DB 1990 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2049
QY 2041 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2100
DB 2050 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2109
QY 2101 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2160
DB 2110 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2169
QY 2161 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2220
DB 2170 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2229
QY 2221 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2280
DB 2230 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2289
QY 2281 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2340
DB 2290 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2349
QY 2341 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2400
DB 2350 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2409
QY 2401 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2460
DB 2410 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2469
QY 2461 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2520
DB 2470 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2529
QY 2521 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2580
DB 2530 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2589
QY 2581 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2640
DB 2590 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2649
QY 2641 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2700
DB 2650 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2709
QY 2701 gaaagttacaagaaccttlaaatalatggaagattatcttcaagaaatagagtt 2760

Db 2710 CCTCAAGACAAAGCCGAGAGCGCTAAACGACAGACACCTCTACCTGACCAACCTTAAT 2769
QY 2761 aatagtttaaatattttgaaacaatattgagcttctggaaaaaacaataatataacca 2820
Db 2770 AACTCTAGTAAGCTGTTTGGAACTCTCTCTCGGGGAAAGATTAAGAACTATCCAA 2829
QY 2821 gaattaaagtcacaaaagtagtgaanaactttaaagaagattataagaagtagtgc 2880
Db 2830 GAACCTATTGACAGAAATGCTCCGAGACTTCTAGCAGAAAGTACTGAAAGACAGCGAC 2889
QY 2881 acattataatgaatcctttaaacaatttgttaaaacttaagaactgaatgaattatca 2940
Db 2890 ACATTCATTAACGAGACCTTCACTACTTCTGTAATCTTAAGCCGATGATTCACACTCT 2949
QY 2941 ttgaatgaatgaatcaaaaggaagaattagaagaagattataatataaaaaaact 3000
Db 2950 CTTAACGATGAATCTAAAGCTGAAGAGCTGAAGAGACATCAATTAAGCTGAAGACA 3009
QY 3001 ttacagttatcatttgattataataataataataataatagaagaattatttgc 3060
Db 3010 CTGCACTAGCTTTCGACCTGTAACCAAGTACAACTGAACTGGAGAGACTTTCGAC 3069
QY 3061 aaaaagaacagtttgytaaatataaalycaaatlaaaaaacttaattataaagaa 3120
Db 3070 AAGAAGAGACAGTGGGCAAGTATTAAGATGCAGATCAAGAAAGTTGACTGCTCAAGGAG 3129
QY 3121 caattgaatcaaatattgaattcaacttaataacccaaagcatgattacaacatttct 3180
Db 3130 CACTTGAAGAGCAACTCAACTACGTAACAAACCAACCACTACTGCGAAGCTCTCA 3189
QY 3181 gttctttaaacaacaaaagaagcttgaatagcagaacactgaacacattagaacaa 3240
Db 3190 GTTCTTTCACAGAAAGAAAGAGAGCGAGATGCGCGAGACAGAAACACTCTGGAGAAC 3249
QY 3241 acaaaaattatttgaacattataaaggacttgaatattataatattgtagaatact 3300
Db 3250 ACCAAGATTCTTCTCAACACACTCAAAAGGCTGTCAAGATTAATTAAGCGAGCTTCT 3309
QY 3301 ccaattaaaactttaaagtaagaatcaatcaaaagaagataatattgcagttagaa 3360
Db 3310 CCTCTAAGACACTCTCTCCAGGAGACATCCAGACGAGATTAATTAAGCGCACTCGAG 3369
QY 3361 aactttaagatttaagtaattagaagaagaattagaagattataatagaagaag 3420
Db 3370 AACTTAAAGGCTCTGTCTAAGCTCGAAGGAACTGAAGGACAACTGAACCTGGAGAG 3429
QY 3421 aaaaaattatcaacttaacagtgatattacattcaattgaactgaatataagaaga 3480
Db 3430 AAGAAGCTCAGGTAACCTCTAGCGGAGATGCATCCTGATCGCGAGCTCAAGAGATC 3489
QY 3481 ataaaaataaaaaattatacaggaattctccaagtgaanaataacaggaatgaacaat 3540
Db 3490 ATTTAAACAAAGAACTAACCCGCAATTAACCCCAAGGAGAAATTAATACACGCTGAATAAC 3549
QY 3541 gcaattagaactttaaacaacaaattctccaagaagaagacagattgttgaacagttgaat 3600
Db 3550 GCACTGGAATCTTACAGAAAGTCTGCGCTGAAGGAAACAGATGTCCTGAGGTCT 3609
QY 3601 gaaagtggatccgacacattagaacaaagtcacacaaagaacacacacacacacatgta 3660
Db 3610 GATCTGGCTCGACACACTGGAGCGTCTCAACCTAAAGAGCTCATCTACTCATGTC 3669
QY 3661 ggaagcagatcttaacacaaatcaacaaatgctcgatgaatgaatgaatgaatgaatga 3720
Db 3670 GGAGCGCATCAATTAATTAACATCTCAAGAACTGAGAAAGTGAAGTGAAGTGAAGTGA 3729
QY 3721 atcatagtaacctatatttgagaatccgaagaagattatgaatgaatgaatgaatgaatga 3780
Db 3730 ATCATGTTGCTTATCTTTCGCGAGAGCGAGGAGGACTTACGATGACCTCGGCCAGGTGTC 3789
QY 3781 aagaagaagcagtaactcttcgtaattgaatgaatgaatgaatgaatgaatgaatgaatga 3840

Db 3790 ACCGTGAGCGTGTCACTCCTCTGATTAATGATTAACATTTCTGCCAAATTCGAGAACGA 3849
QY 3841 tatgaagtttattatttaaaccctttagcagtggtttatagaagtttaaaaaacaata 3900
Db 3850 TACGAAGTCTCTATCTGAACCTCTGGCAGGCGTGTATAGCTCTCAAGAAACAGCTG 3909
QY 3901 gaaataacgtaataatgaattatgtaattgttaagaagatttaaatcaagatttaatt 3960
Db 3910 GAAATTAACGTGTGACCTTCAATGTCMAAGTGAAGACATTTTGAAAGCCCTTTAAT 3969
QY 3961 aaacgtgaacatttcaaaaatgltttagaatcagatttaattccatataagaattlaaca 4020
Db 3970 AAGAGAAATTTCAAGAACGCTTGGAGAGGACTGATTGATTCCTAATAAGACCTGAGC 4029
QY 4021 tcaagtaattatgttgcacaaagatccataaaatttctttaaataagaagaagaatga 4080
Db 4030 TCTCTTAACCTGTTTCAAGGACCCATCAAGTCTCAATTAAGAGAGAGATTAATA 4089
QY 4081 ttcttaagcagttatattatattatagaattcaagaattcagaatgaatgaatgaatga 4140
Db 4090 TTTCTGTATGTTACAACTATATCAAGGACTCCATGACACGATATCAATTTTCTTAAT 4149
QY 4141 gatgtcttgyatatataataatataatccgaaaaatataaatcagaatttagatcaatt 4200
Db 4150 GATGTGCTGGGTATTACAAGATCTTGAGGAAATTAACAGTCTGACCTGACCTTATT 4209
QY 4201 aaaaaatatcaacgaacaaagaagtgaaatggaataacttcccttlltaacaat 4260
Db 4210 AAAAGTATATCAACGATTAAGCAAGCGAGAAATGAATAATCTGCTTCCGTAATAC 4269
QY 4261 attgagaccttatataaaacagtttaaagtaattgaatttlltgaattatga 4320
Db 4270 ATGGAACCTCTGACAAAGACAGTGAACGACAAATGACCTCTTCTTAATTAACCTGGAG 4329
QY 4321 gcaaaagtttcaaatatatacatatagaagaatcaaacgtagaagttlaaaaataagaact 4380
Db 4330 GCCAAGTCTCTCAACTATTAATCTTACGAGAGAGCAATGTGAAGTTAAATCAAGAGCTG 4389
QY 4381 aattacttaaaaaatllcaagaacaaattggcagattttaaanaaaatacaaatgcgt 4440
Db 4390 AACTTACTCAAAACATCAAGCAAGCAAGCTGCGAGATTTCAAGAAATTAATTAATTCGTC 4449
QY 4441 ggaattgcgatttatacaacagattataacataataacttaacttgaacaaagttccat 4500
Db 4450 GGAATTCAGACCTGTCTACCGATTAATTAACCAACAAATCTCTGCAACAAATTTCTGTC 4509
QY 4501 acagatgattgttttgaacaaacttgcataaacggttttataacttaacttgaatgaac 4560
Db 4510 ACTGGCATGCTGTGCAAAACCTCGCCAAACAGTCTGAGCAATGCTGCCAGGCAAC 4569
QY 4561 ttgcaagttatgtlaacatttcaacaacaaatgcaatgcaaaaaacaatgctccacaat 4620
Db 4570 CTGCAAGGATGCTGAAACATCTCCAGCACCAATGCTGAAGAAACAGTCCGCCAGAT 4629
QY 4621 tctggaatttgaacatttagatgaagaagaagatgtaaatggtttatataatcaaa 4680
Db 4630 AGCGCTGTTTGAAGCATGTGACGAGCGGAAAGATGAAGTGTCTCTGAACTACAA 4689
QY 4681 caagaagtgataaattgttgaanaatccaaatccacttltgaacgaanaataatgtga 4740
Db 4690 CAAGAAAGAAATTAAGTGGTGGAAGAACCCAAACCTTACTGCAATGAATAAACAATGGCGG 4749
QY 4741 tgtatgcagatgccaatgtaacgaagaagattcaagtagcaacggaagaanaatcaca 4800
Db 4750 TGTGACGCCGATGCTAAATGACCGAGAAAGACAGGCGCTTAACGGAAGAAATTAACA 4809
QY 4801 tgtgaatgaataaacttaattcttaacttlltgcagatgtatttlltgcagttcct 4860
Db 4810 TGGAGTGTACTTAAGCCGACTCTATCTCACTCTGAGAGGAGATTTTGTCTGACGCTT 4869
QY 4861 aacttcttagaataatcaacttcttaataaactcaatgaatgaatgaatgaatgaatgaatga 4920
Db 4870 AATTTCTGGGATCTCTCTGCTGATCTCTATGCTATCTGTAACGCTTCACTTA 4929

RESULT 9
LOCUS CVEJ31294 4940 bp DNA SYN 24-FEB-1999
DEFINITION Cloning vector pBSK+-MSP-1s/FCB.
ACCESSION AJ131294
VERSION AJ131294.1 GI:4455908
KEYWORDS Cloning vector: major surface protein-1; msp-1 gene.
SOURCE Cloning vector pBSK+-MSP-1s/FCB.
ORGANISM Artificial sequence: vectors.
REFERENCE 1 (bases 1 to 4940)
AUTHORS Bujard, H.
JOURNAL Direct Submision
TITLE Submitted (14-DEC-1998) Bujard H., ZMBH - Zentrum fuer Molekulare
Biologie, Universitaet Heidelberg, Im Neuenheimer Feld 282, 69120
Heidelberg, BW, GERMANY
2 (bases 1 to 4940)
Pan, W., Rayot, E., Toille, R., Frank, R., Mosbach, R., Turbachova, I. and
Bujard, H.
REFERENCE
TITLE Vaccine candidate MSP-1 from Plasmodium falciparum: a redesigned
4917 bp polynucleotide enables synthesis and isolation of
full-length protein from Escherichia coli and mammalian cells
Nucleic Acids Res. 27 (4), 1094-1103 (1999)
99128299
FEATURES
source
1..4940
/organism="Cloning vector pBSK+-MSP-1s/FCB"
/db_xref="taxon:85834"
/focus
1..4940
/organism="Plasmodium falciparum"
/strain="FCB-1"
/db_xref="taxon:5833"
10..4929
/gene="msp-1"
10..69
/gene="msp-1"
10..4929
/gene="msp-1"
/function="surface protein"
/codon_start=1
/evidence=experimental
/product="major surface protein-1"
/protein_id="CA36903.1"
/db_xref="GI:4455909"
/translation="MKIFFLCSFLFITITQCVTIHESIYQELVKRIKLEDAVLGTS
LFQKEMVLNEGISTAVITTPGSKGVSAGSGSVASGVSAGSVASGVSAGSVAS
GGSGNSRRTPNSDSDAKSYADLKHVRNLTITKELKPOLDELNHNMLDVK
HGFKLIDGEEINELLYKNFYFDLRAKLVNDVCANFYCOIPMLKIRANLDELK
KIVFGYRKPLDINKONVGMEDYIKKKKTIENINILIEESKTIIDKKNAKPEERK
KLYQAYDSTYINKOLEBAHNLISYERKIDITLKKNENIKELDKINEKNPPRANG
NTPNLTLDKNKKIEHEKEIKETAKTIRKINISLFDPLELEYLEKKNKIDISAKV
EKKESTENPEYPGVTPYPLSYNDINNALNELNSFGDILPDPYTERKKNITDNRK
KINIEIKERIKIEKKIESDKSYEDRSRLNDITTEYKLEINLETYKFNNDILTN
PEKMGKRSYKVEKLTHTNTFASYSNHNLEKLIKAKVMEDYSLRNIVEKELKY
YKNLSKINENIETLVENIKKDEEOLFEKKITKDEMKPDEKILEVSDIKYOVQKYL
MKNIDELKKTQULIKVNLKHNHVNASTKOEKOPYLLIYKLEIDKLKMPKVE
SLINEKKNIKTEGSDNSEPTGEITQOATTPQOQGSALGSGVQAQOEQOA
OPPEVYPPEAKAOPVPPAPVNNKTEVNSKLYLEKLEFLNTSYCHRYLTHSHT
NMEEKIKOYKIRKEESKLSGCDLILFNINNIIPVMTSMFSLNNSLSQLEMEIYE
KEWVCNLYKTKDNKDKIKNLLEAKKVSSTVKTLSGSMGLSTTPDRKPEVANDOTS
HSTNINNSLKLEFNISLGKKNKIYQELIGOKSSSEMYEKIKLDSTPTFNESTNRYK
SKADINLSLNSDKRRKLEEDINKLKTITQSLDLYNKYKLEIRLFPKRTKVKYKIM
QIKKTLLEOLESKLNSLNPNKAVHLONSVFNNKKEKEIAETEMTNTILIKHY
KCLVYKNGESSPLKTLSEESLQTEDNVSLENFYKLSLEKLNKLNLEKKISYL
SSGLHHLIAELKEVINKNVTGNSPENNVDVNNALSESYPKFLPCTDVAATVASEGS
DTLEOSOPKPPASTHGASNTITTSQNDDEVDVITVPIGSESEDDYDILGOVYGT
EAVTPSYDINILISKIENEYEVILKLAAYISLKLQLENNVMTFVANKDILNNGFN
KRENENVNLESDLIPYKDLTSSNYVVDKYEKLNKRKRPFLUSVYIKDSIDTDINF
ADVLDGYKILSEYKSDSDSIKKYINDQGENEKLPLNNIETLYKTKVNDKIDLEFV
IHLKAVLVNYTEKSNVEVKKIKELNLYKTIODKLADFKKNNNFVGIADLSTDYNNHNL

BASE COUNT 1669 a 1156 c 1060 g 1055 t
ORIGIN
LTKPLSTGWNVEENLAKTVLSNLLDGNLOGMLNISOHQVKKKOCPPONSGCFRHLDEREE
CKCLNLYKOEGRKCVENPPTCENNGGCDADACKEEDSGSNGKRTICECKRPSYP
LFDGIEGSSNFIPIFLILMLIYSFI"
Query Match 57.2%; Score 2812.8; DB 56; Length 4940;
Best local Similarity 73.2%; Pred. No. 7.5e-253;
Matches 3603; Conservative 0; Mismatches 1317; Indels 0; Gaps 0;
QY 1 atgaagatcatcttcctttagtgcattcttctttttattataatcacaaatgta 60
DB 10 ATGAAATCATTTTCTTCCTCTCTCATTTCTGTTTATCATCAATCACTAGTCGCTG 69
QY 61 accatgaagaagtataaagaactgtcacaataactagaagcttagaagatgagatgt 120
DB 70 ACCCAGGAATCCTATCAGGAGCTGTTAAGAACTGGAAGCTTGAAGATGCCGTCTT 129
QY 121 accagttatagttattctcaaaagaaataatgtataatagaagaacaatgtaaga 180
DB 130 ACCGGATACAGCTGTTCCAGAAAGAGAGATGTGCTGATGATGAAGAGCATGCGACG 189
QY 181 gctgtacaactagtaacacctgtgtcaaaaggttcagttcaggtcaggtgtcaggtg 240
DB 190 GCCGTTACAAACACACACACACCGCTGCTAAAGGCTGTGCTGACGCGTGGCTCGG 249
QY 241 tcaagttcaggtcaggtcaggtcaggtcaggtcaggtcaggtcaggtcaggtcaggt 300
DB 250 TCTGTGGCTCTTGCGGCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 309
QY 301 gcttcaggtcaggtcaggtcaggtcaggtcaggtcaggtcaggtcaggtcaggtcag 360
DB 310 GCAGAGGCGCGGTTCCGGGAACACTGCAAGAACCAATTCATGCAACACTGACGATTC 369
QY 361 gatgtcaaatcttcaggtcaggtcaggtcaggtcaggtcaggtcaggtcaggtcag 420
DB 370 GACGCCAAGTCTTACCGCCGACCTCAAGCACCGAGTAGAAGAACTATCTCCATCA 429
QY 421 gaactcaaatatctcaactcttgaattgaattgaattgaattgaattgaattgaatt 480
DB 430 GACCTGAAGTACCCACAGTTGTTGACCTGACCTGACCTGACCTGACCTGACCTGAC 489
QY 481 atccatggttcaaatatctgaattgaattgaattgaattgaattgaattgaattga 540
DB 490 ATTCAATGCGCTCAATATATGATGATGAGGTTAGCAAGATGCAATGACCTGATCA 549
QY 541 ttaacttattctgatttataagaagcaaaataaagtatgtagtgcattgattat 600
DB 550 TTGAATTTCTACTTTCGACCTTGTGAAGGCCCAACTGATGACCTTTCGCCAATGA 609
QY 601 tctcaaatccttcaatcttaaatcttcgagcaaatgaattagagctacttaaaact 660
DB 610 TGTCAATTCATTCATTTGAAGATGACAGCCACAGAGTGGACCTATTTGAAGAAG 669
QY 661 gtttcgatatagaaaacccatagaacaattaaagaataatagaagaatagaagaat 720
DB 670 GTCTTGGAATGCGAAGCCTCTGCACAAATCAAGACAAATGTGGGAAGAAGAGAT 729
QY 721 tacattaaaaaaataaaaaaccatagaataataatgaattaaatagaagaagaat 780
DB 730 TATATTAAAAAAGATTAAGAACCATTCGAGAACATTAAACAGCTGATCGAATCCAA 789
QY 781 aaaaacattagataaaaataagaatgaactcaaaagaagaagaagaagaataatacc 840
DB 790 AAGACCATAGCAAAAAATTAAGATGCAACCAAGAGAGAGAAAAAGAAAGATGTG 849
QY 841 gttcaatatgaattcttctattcaataaacaattagaagaagcacataatgaaga 900
DB 850 GCCCAGTAGCACTGTTCATCTTAACAAACACCTTGAAAGAACCCATTAACCTCAT 909
QY 901 gttttagaaaaagtaattgacactttaaaaaaaatgaagaacttaagtaattgatt 960

Db 910 GTACTGAGAACGCGATAGACACCCCTCAAGAAATGAAATATATCAAGAACTGCTGCAC 969
Qy 961 aagataataagaaatlaaaatccccccacgccaatttcggaaatacaccaaatctctc 1020
Db 970 AAGATTAAATGAATTAAGATCCTCCGACGCAACTCTGGGAACCCCTTAACACGCTG 1029
Qy 1021 ctgtgaagaacaaaacacgcgagacacgaaagaaatlaaagaatttgcgaact 1080
Db 1030 CTGGACAAACACAGAGATAGAGAGACGACGAAAGAGATCAAGAGATCCCAAAACC 1089
Qy 1081 attaaattlaataatgatagttatttactgacacattgaattgaatactattaa 1140
Db 1090 ATTAACTTCAACATGATCTCTCTTACTGATCCCTTGAGCTGAGTACTGCTTGAG 1149
Qy 1141 gaaaaaataaataatgataagtcgaaggttgaaacaaaggatccaactgaacc 1200
Db 1150 GAGAAATTAATTAACATCTCCGCCAAGTGCAGCAAAAGAAATCAACCGAACCT 1209
Qy 1201 aatgaataccaatgaggtacttaactcttgtaataagatattaacaatgcttca 1260
Db 1210 AATGAATATCCCAATGCTGTGACGTACCTCTGTTATTAACATTTCAACAACTCTTC 1269
Qy 1261 aatgaacttaattcttcttggtgataatlaalccaattgattatacaaaagacaa 1320
Db 1270 AACGACCTCAATAGTTCGTGACTGATTAAACCCCTTGATTAATGGAAGAACCTCT 1329
Qy 1321 aaaaacataatactgataatgaagaaaaaattcaataaagaaattgaagaaaaat 1380
Db 1330 AAGAAATTTTACACACACATGAGAAAGAAAGTTATCAACGAAATCAAGAGAAATC 1389
Qy 1381 aaaaatgaaaaaaaaatgaaatcgtataaaaaatctgaagaacagatcaagctc 1440
Db 1390 AAAATTGAAAGAAAGAAATTTGAGAGTGCACAGAAAGTTACCAAGACCCGACAAAGT 1449
Qy 1441 ttaaatgataaacaagaataatgaagaaaaatctactaaatgaattatgataagcaatc 1500
Db 1450 CTAAAGATATCATCTAAAGATGATAAGAAAGCTGAAGAGATCTATGATTTCCAATTC 1509
Qy 1501 aataataataatgatttaactaatttcgaaaaaaatgaggttaagaatactcataaa 1560
Db 1510 AACAAATTAATGACCTGACCAACTTCGAGAAATGATGGAAAGGTAATCTTCAAA 1569
Qy 1561 gtgtgaataactacacacataataactttctgaactcttaagaatacttaaaatctc 1620
Db 1570 GTGGAGAAACCTGACACACATATACCTTTGCATCTATGAGAAATTTAAGCAATCTT 1629
Qy 1621 gaaaagttaacaagaagcttlaaataatgagaagatlaattcttaagaataatagtagt 1680
Db 1630 GAGAAAGCTCACCAAGCTCTTAAGTATGAGAGATATCTCTGCGAACAATGTTGTG 1689
Qy 1681 gaaaaagaattaaaatataataataataataagaacaaatagaaatgagatgaaca 1740
Db 1690 GAGAAAGAACTAAAGATTATCAAGAAATCTCAATGAATCGAAAAACGAGATCGAGACG 1749
Qy 1741 ttaagtgaataatlaaataagaatgagaacagcttctgaaaaaaaatcttaagaac 1800
Db 1750 CTGTTTGAACATTAATGAAGATGAAGACACTGTTTGAAGAGAAAGATTAACAAAGAC 1809
Qy 1801 gaaaaataaacgcgaatgaagaaatcttgaagatatactgacatgtaaaatgacaagtcca 1860
Db 1810 GAAATTAATCAACGATGAGAAATCTCTGAGAGTCTCGATATTGTAAAGTCCAAAGTGACG 1869
Qy 1861 aagaattttatgaacaaaaatgacgaatttaaaaaagactcaatgtatttaaaaaat 1920
Db 1870 AAGGTCTCTCTGATGACAAAGATTGATGAACTCAGAAAGACTCAATCTTCTGAAGAAC 1929
Qy 1921 tgaagaatlaaacaataatataatgttcccaattcttaacaacaagaagaaatgaagaa 1980
Db 1930 GTGGAGTTAAACATTAATTAATGATGTCGCCGAATAGTTTAAGAGAGAGATTAAGACGAGAA 1989
Qy 1981 cctattataatgtagtctgaaaaaaagaattgataaattlaaagtgctcatgctaaag 2040
Db 1990 CCATATACCTGATCTGATCAAGAAAGATTAAGCAAACTGAAAGTGTTCATGCCCAA 2049

Qy 2041 gtaagaatcatgtaataatgaagaaaaaaacataaaaaacgaaggttcaatcgataat 2100
Db 2050 GTGAGAGACCTGATTAACAGAGAGAAAGAACATTTAAACTGAGAGACGTAGATTAC 2109
Qy 2101 tcggaacatcaaacgaaagagaaataacagagacaagcaactcaaaaacttgacacaa 2160
Db 2110 TCCGAGCCTTCCACAGAGAGAGATTAACCGGACAGCTACCAACAGCCGGACACAG 2169
Qy 2161 gcagagcttgcttagaagagatctcaglaacaagcacaagacaaagaacaaagca 2220
Db 2170 GCCGTTTACGCTTCGAAGGCGATAGCTGCAAGCTCAAGACACAAAGCAGAGCGCA 2229
Qy 2221 caacacccagtaaccagtaaccagtaacgaagaagcaagtcaccacacacagaa 2280
Db 2230 CAGCTTCAGTGCAGTGCCTGTTCCAGAGGCTAAAGCTCAAGTGCCTAACCCAGCT 2289
Qy 2281 ccagtaataataaactgaanaatgcttccaaatagattatcttgaanaatataatga 2340
Db 2290 CTTGTAATTAACAGACCGAGATGTCAGCAACTGACTACTTGAGAAAGCTTATGAG 2349
Qy 2341 tttaataacttcaatataatgtaacaataatatttggtttcacactcaactgaac 2400
Db 2350 TTCTGAATTAATCTTAATCTGCAAAATATATCTGCTCTGTCACAGCACTATGAAC 2409
Qy 2401 gaaaagataatlaaacaataaataatlaacaagaaggaagaagaagaatlaagtlc 2460
Db 2410 GAGAAATTTTAAACAGTAACAGATTAACCAAGAGAGAGATTAACCTCTCTTGT 2469
Qy 2461 gatccatagacttattgtttaatacaaaataacatctcttaatgtaatctatgct 2520
Db 2470 GATCCACTGAGCTGCTGTTCAATATCCAGAACAACTTCCGTATGTTATTTGTTTC 2529
Qy 2521 gatagctlaaaccaatgattatcaacaactatlaagaaatttgaagaaagaatgct 2580
Db 2530 GATAGGCTCAACAAATCTCTCTCAACTGTTTCATGAGATATATGAGAAAGATGATC 2589
Qy 2581 tgaatttataataactlaagatagaaatgaagaaatlaaataatlaagagagagaa 2640
Db 2590 TGCACCTGATTAATCACTCAAGAACGACAAAGATTAAAGACCTTCTGAGAGAACTAAG 2649
Qy 2641 aagaatccacactgtlaaaaactcttcaagttcaatgtaacgaacatataatca 2700
Db 2650 AAGTCTCCACCTCTTTAAACTCTCTCTCCAGCTCCATGCAACACACTGCTCACA 2709
Qy 2701 cctcagataaacgcgaagtaagtgaatgataatcacatctcaaaaattgaa 2760
Db 2710 CCTCAAGACAAAGCCGAAAGTGAAGCTTAACGACGACACTCTCACTCGAACCAACTTAAT 2769
Qy 2761 aatagtttaaatatttgaaaaacataatgagcttggaaaaaaacaaaatatataccaa 2820
Db 2770 AACTCACTGAACCTGTTTGAAGACATCTGCTCTGCGCAAGTAAGAACATCTACCA 2829
Qy 2821 gaattatagtcacaaaagtagtgaanaacttlaagaaagatctlaaagaatgtagt 2880
Db 2830 GAACATTATGAGAGAAATGCTCCGAACTTCTAGAGAGAAATTAAGTGAAGACGAGAC 2889
Qy 2881 aacttataatgaatcttcaaaaatttgaanaatcctaaagctgaatgataatca 2940
Db 2890 ACATTTATTAACGAGAGCTTCACTTAATCTGTAATCTTAAGCCGATATCAACTCT 2949
Qy 2941 tgaatgaatgaatcaaaaaggaagaatgaagaagatataatlaaanaaaact 3000
Db 2950 CTTAACGATTAATCTTAACCTTAAGAGCTGAGAGAGACATTAATTAAGTGAAGAGACA 3009
Qy 3001 ttaacatcatcttgattatataataataataataataatgaagaatctatgtagt 3060
Db 3010 CTGCAACTGAGCTTGCAGCTGTCACAAAGTAACTGAACCTGAGAGAGACTCTGAC 3069
Qy 3061 aaaaagaagaagtggttaataataatgaatgaatgaatgaatgaatgaatgaatgaatga 3120
Db 3070 AAGAGAGAGAGAGAGAGAGATTAAGATGAGATCAAGAAAGTGTGCTCAAGAGAG 3129

OY	3121	catltaagatcaaaatttgaattcaacttaataaccacaaagcatggtttcaaaactttt	3189
Db	3130	cagcttgaaacaaactcaactcactbacaatccgaacacgtactgcagacttctca	3189
OY	3181	gtttcttcaacaaaaaagaagcttgaattagcgaacgtgaanaacacatlagaaac	3240
Db	3180	gtgtttcttcaacaaagaaagaaagccagatgcgcgacagaaacactcttgagAAC	3249
OY	3241	acaaaatattatttgaacattataaagaacttglttaattataattatgltgaatcatc	3300
Db	3250	accAAGATTCTTCTCAAAACACTCAAAAGGCGTCGTACGATTATTAATGGCAGCTTCT	3309
OY	3301	ccattaaaaccttaagttgaagaatcaattccaagaagaagatlaattatgccaagttlagaa	3360
Db	3310	cctctgaagactctctccgagagagacatccAACCGAGATTAACGCCAGCCTCGAG	3369
OY	3361	aactttaaagctatlaagtlaaattagaaggaaatlaaagatlaattlaaattagaag	3420
Db	3370	AACTTCAAGGTCGTGCTTAAGCTCGAAGGACACTTAAGACACTGAACTGAGAGAG	3429
OY	3421	aaaaaatlcatctatccaaagttgaatcatcalttaatttgcttgaattlaaagaagta	3480
Db	3430	AAGAAGCTAGACTCCTCTTACGGGAGCTGCATCAGCTGATGCGGAGCTCAAGAAATC	3489
OY	3481	ataaaaataaaaaattataagttatttccaaagtgaanaataaataagatglttaacat	3540
Db	3490	ATTTAAGAACAAGACTACCGGCAATAGCCCAAGCGAATAATTAACAGACTGATATAC	3549
OY	3541	gcattgaatcttcaaaaaaatttccccaagaagaagatglttgaacagttgtaagt	3600
Db	3550	GCAGTGCATCTTTCAGAAAGTTCTGCTCGAAGGAACAGATGTGCCACTGTGGTCT	3609
OY	3601	gaaaagtgaatccgcacatatagaacaagaatcaaccaagaanaaccagatcaactatgta	3660
Db	3610	GAATCTGGCTCCGACACACTGAGGACAGTGTCAACCTTAAGAACCTGACATCTCATGTGC	3669
OY	3661	ggagcagagcttcaacacaataacaacatacaaaaatgctgatatgaagttagatgaagta	3720
Db	3670	GGAGCCGGTGCTCAATTAACATTAACCATCTCAACAGCTGAGAGTAGGTGCTATGCTGC	3729
OY	3721	atcctataccataatttggagaatccggaagaagatctatgcatgtttagaagaagtagta	3780
Db	3730	ATCATTTGGCTATCTTCGGCGAGAGCGAAGAGACTACGATGACTCGCCAGTGGTGC	3789
OY	3781	acaggagaagcagtaactctctccgtlaattgtataacatlaacttcttaaatlgaanaatgaa	3840
Db	3790	ACCGGTGAGGCTGCACTCTTCGCGATGTGAATCAATCTGTCCAAATGAGAGAACAA	3849
OY	3841	tatgaagtttatattataaacccttaagcagagtglttatagaagttcaaaaaacaatta	3900
Db	3850	TACAAAGTGTCTATTTGAAACACTCTGGCAGGGGTCTATTAAGTCTCTCAAGAACAGCTG	3909
OY	3901	gaaaataagctatgacatttaagttaagttaagtattttaaacttcaacgtattat	3960
Db	3910	GAGAAATTAACGTGATGACCTTCAATGTCAACGTGAAGAGATTCTTAACAAGCGCTTTAAT	3969
OY	3961	aaacgtlgaanaatttcaaaaaatgltttagaatcaagattlaattccatatagaattlaaca	4020
Db	3970	AAGAGAGAAATTTCAAGAACGCTTGGAGAGAGACTTGATTCCTATTAAGACACTGACC	4029
OY	4021	tcaagtaattatglttgcataagaatccatataaatttcttaataaagaanaagaataaa	4080
Db	4030	TTCTCTTAACACGTTGTCAAGGAGCCCATTAAGATTCTCAATTAAGAAAGAGGGATTAA	4089
OY	4081	tctcttaagcagttatataattatlaagagttcaalagatlaagatataaattttggcaat	4140
Db	4090	TTTCTGTCTATGTTTACAATAATTAACAAGGACTCCATGACACCGGATATCAATTTGCTAAT	4149
OY	4141	gatgttcttggatataataaaattatccggaanaataaatacagatttgaatttcaat	4200
Db	4150	GATGTGCTGGGGATTATTAACAAGATCCGTGAGCGAAAAATTAACAATCTGACACTGACTATTT	4209
OY	4201	aaaaaatatatacgaacaaacaagtlgaanaatlgagaatatcccttccttttlaaacat	4260

Db	4210	AAAAATATATACACGATTAAGCAGGCGGAGATGAAAAATATCTGCCCTTCTCGAATTAAAC	4269
Oy	4261	attgagacctatataaacaagttaa tga taaattgatttalttg ttaattcatattagaa	4320
Db	4270	ATCGAAGCCCTGTACAGAGACGTGAACGACAAAAATCGACCTCTCTGTAAATTCACCTGGAG	4329
Oy	4321	gcataagttcttaattatatacatatgagaaatcaaacgtlagaagtgttaaaataaagaactt	4380
Db	4330	GGCAAGTCTCTACACTATCTTACGAGAGAGCAAAATGTGGAAATGTAATAATCAAGGAGCTG	4389
Oy	4381	aattacttaaaacaattcaagaacaatttgcagattttaaataaaataacaatttcgtt	4440
Db	4390	AACTACTCTAAAAACAATCCAGACAAAGCTGGCAGATTTCAAGAAAAATATACATTTTCGTC	4449
Oy	4441	ggaattgctgattatlcacaacagattlaaacataaacttalttgcacaaagttccttaag	4500
Db	4450	GGAAATGCAGACCTGTCATCCGATATATACCAACAACTCTCTGACCAAGTTTCTGTC	4509
Oy	4501	acaggtatagtttttggaaaaactctgtctaacaacgltttcttcaattacttbatgtgaaac	4560
Db	4510	ACTGGGATGTGTTCGAAAACTCGCCAAACAGTCTCAGCAATCTGTCACGGCAAC	4569
Oy	4561	ttgcgaagttatgttaaactttcaacaacacagatggttaaataaacaatggtcacacaat	4620
Db	4570	CTGCAGGCGATGCTGAACTTCTCCGACACCAATGCGTCAAGCAAAACAGTGCCTCCAGAAT	4629
Oy	4621	tcbyga tgtttcaagacatttagatgaaagaagaaatgtanaa tghttalttaattaca	4680
Db	4630	ACGCGCTGTTTCAGGCATCTGGACGAGCGCAAGAAAGTGCAGTCTCTGCAACTCAAA	4689
Oy	4681	caagaagtgatataatgttgttgaanaatccaatccactacttgaatogaanaataatggttga	4740
Db	4690	CAGAGAGGAGATTAAGTCTGGAGAACCCAAACCCATCTGGAATGTAAACAAATGCGGGG	4749
Oy	4741	tgtgatgcatgtccaaatgtaaccggaagaagattcaagtgacaacggaagaanaatcaca	4800
Db	4750	TGTGAGCCGATGCTTAAATGCAACCGAGGAAGACAGCGGCTCTTAACGGAAGAAAAATCACA	4809
Oy	4801	tgtgaatgactaaacctgattcttataccacttttcga tggatatttctgcagttcctct	4860
Db	4810	TTCGAGTGTACTTAAGCCGACACTTCCTATCCACTTCGACGGGATTTTTCCTCCACACTCT	4869
Oy	4861	aactcttagagatatcaattctattaaactcaacatgtaataatatataacagtttaactaa	4920
Db	4870	AAATTCTCGGGAATCTTCTCTGCTGATGCTCATGCTATGCTCTGTACGGTCACTCTTA	4929

RESULT 10
AF218248
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
Source

AF218248 5085 bp DNA INV 24-JAN-2000
Plasmodium falciparum major merozoite surface antigen (gp195) gene,
complete cds.
AF218248
AF218248.1 GI:6740099

malaria parasite *P. falciparum*.
Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 5085)
Shan,Z.X., Yu,X.B., Li,X.R., Ma,C.L. and Fang,J.M.
Molecular cloning and sequence analysis of major merozoite surface
antigen(gp195)gene of Plasmodium falciparum isolate FCC1/HN
Unpublished
2 (bases 1 to 5085)
Shan,Z.X.
Direct Submission
Submitted (21-DEC-1999) Department of Parasitology, Sun Yat-sen
University of Medical Sciences, 74 Zhongshan Road II, Guangzhou,
Guangdong 510089, P.R.China
location/Qualifiers
1..5085
/organism="Plasmodium falciparum"

```
/isolate="FC01/HN"
/db_xref="taxon:5833"
/country="China"
1..5085
/gene="gp195"
1..5085
/gene="gp195"
/codon_start=1
/product="major morozoite surface antigen"
/protein_id="AA027526.1"
/db_xref="gi:6740100"
/translation="MKILIFLCSPLFIINTQCVTHESYOELKKLEALDELAVLNGYS
LEOKEMVLENGTSGTAVTSTPSGSGVSGSASVASVSGSGSVASGSGNS
RRTNPSDSDSAKSVADLKHRYONYLTETKEIKLPEPLDINMLTCDIHPFGY
LIDGEEINELLYKNTFYDILRAKLNACANSYQIFPNLKRANEDLVKIKYFGY
KRLDNIKDNYGKMDYTKNKTITANIINELLESKKTIDONKANDEGKKLYQAO
YLFYLNQLOEAHNLISLEKRLIDTLKNEIKMLLEDIDKIKDAKRLTGSFNP
LPEKKKEVEGHEEKIKELIANTIKENIDSLFDPLELEYLREKKKYDVTSPDPT
KSVQIPKVPNGIYVPLPLDINSLAADNDKNSYGLMDNDTEKINEKIITDNKE
KRFINNIKQIDLEEKNIHNTKEQNKLLIDYEKSKDYDEILEKFEYEMKRNMPDK
DVNDKIFSAKRTYVWEKOKYNNKSSNSNYNOKLKAISYLEIDYSLRKISKEDF
NHYYTLKTGLADIKKLTLEIKSEKLEKFKGLTHSANALESVDIKLOVKVL
LIKILEDRLKLELFLKNAQLDSHVPNIYKPNKPEYVILVLEKVEDKLEFIPKY
KDMLKKEQAVLSITQPLVAASETDEGSHHTLSQGEETEETVGHHTVITL
PPTQSPPEKVVVENSIEHKSNDNSQALITVYLKLDIEFLTKGYICHKYLIVSNS
MDOKLEYNLTPREKNELKSDPLDILFNIONNTPAMYSILDSNNNDLOHLPFLYO
KEMYYLKLKEENIKKLEOKOITGTSSTSSGANTVNTAOSTNSNSQNSNA
SSTNQNCAVSSGPAVVEESHDLPLVLSISNDLGIYSLNLGKRTVPNPLTISTT
DEMEFENILKNDYFENDIKQFVKSNSKYITGLTQKNALEIKLKDITLQSLF
DLUNYKILKDLRNFKKKELQODKQIKLTLKLEBLSKNSLNPNHNLQNFVVF
NKKEAEIAETENTLENTKILKHVKGLVYNGESSPLKTSEVSIOTEDYANLEK
FRVLSKIDKLDNLHLGKKKLSFSLSGHLITELKVKIKKNTGSPSNKKVN
EALKSYENLEPEAKVTVTPPODPVPSPLSVRVSQSSGSKTEQIPTSLLTEL
QOYVQSOYDEEDSVLVPFPGSESDNDIDQVTEBAISVTDNLDSGENEDY
ILPLAGVYRSLKQILEKNITTENLNDLNSLKRKRFLVLEDSLQOKFISS
NEVIIEDESKLINSBOKNTLSKYKIESEVENDIKFAOEGISYLAQKIDLES
IKVYIEKEKEPSPPTPPSPAKTDEQKSKFLPFLNIETLYNNLVNKLDIVLIN
LKAKINDCNVEKDEAHVKITKLSDLKIDDKIDKRNDEPAIKLINDPKRMGLG
KLSTGIYONPNITISKILGKIPDMKINISIOHOCVKNCPENSCFPHLDERECKC
LLNKQESDKCYENPNPTCNENNGCCDADACTEEDSSNGKLTCECTKPDSTPLFD
GIFSSSNFELISFLILMLILYSFI"
BASE COUNT 2244 a 671 c 678 g 1492 t
ORIGIN
```

Query Match 56.2%; Score 2766.8; DB 5; Length 5085;
Best Local Similarity 74.2%; Pred. No. 1.4e-248;
Matches 3800; Conservative 0; Mismatches 1087; Indels 231; Gaps 15;

```
1 atgaagatcatatctcttattatgctcatctctttttttatataaatacacaaagtota 60
1 ATGAAGATCATATCTCTTTATTATGTTTCATTTCTTTTATTATAAATACACAACTGTGA 60
61 acacatgaagatcatcaagaactctgcaaaaaaactagaagctttagaagatgcagatcg 120
61 ACACATGAAGATCATCAAGAACTGTCAAAAACTAGAAGCTTTAGAAGATGCAGATTCG 120
121 acaaggttatagttatttcaaaaggaataatggtataatgaaggaagaagtggaaca 180
121 ACAAGGTATAGTTATTTCAAAAAGGAATAATGATTAATGAAGAACGAAGTGAACA 180
181 gctgtacacagtagcacacctggttcaaaaggtttagtgcaggtgcaggtgcagtgcc 240
181 GCTGTACACAGTAGCACACCTGGTTCAGGTTCAGTTACTTCAGGTG----- 230
241 tcaagttgctcagtggtcagcttgctcaggctgcaggtgcaggtgcaggtgcagtt 300
241 TCAAGTTGCTCAGTGGTCAAGTCTTCAGTTGCTTCAGGTGCTTCAGGTGCTTCAGTT 282
301 gcttcaggtggttcaggttaattcaagaagtagacaatccttcagataaattcaagtatca 360
301 GCTTCAGGTGTTCAAGGTAAATTCAGAGACCTTCAGATAAATTCAGGTGATTCAGTCA 342
```

```
361 gatgctaatacttagctgattttaaaacacagagtagaagaattacttgtaactatcaaa 420
361 GATGCTAATACCTTAGCTGATTTTAAACACAGAGTAGAAGAATTACTTGTACTATTAA 402
421 gaactcaaatatcctcaactccttgatttcaactaatcatatgtaactctgtgtgataat 480
421 GAACCTCAAAATATCCTCAACTCCTTGATTTCAACTAATCATATGTTACTTGTGTGTAAT 462
481 attcaatgcttcaaatatcaattatgtagatagaagaataatgaattatataaa 540
481 ATTCATGCTTTCAAAATATCAATTTGATGATATGAAGAAATTAATGAATTAATTA 522
541 ttaaatcttatttgattatataagaagaaaataatgaatgtagatgtagtctat 600
541 TTAACCTTTATTATGATTATTAAGAGCCAAATTAATGAATGATGATGATGATGAT 582
601 tctcaaatccttcaactatccttaaaatctgctgcaatgtaattgagcttacttaaaact 660
601 TCTCAAACTCCTTCAACTATCCTTAAATCTGCTGCAATGTAATTGAGCTTACTTAA 642
661 gttctcgatagatagaaccatagacaatataatgaataatgtagaagaatgtagat 720
661 GTTCTCGATAGATGAACCATAGACAATATAATGAAGATAATGTAAGAAATGGAAGAT 702
721 tacattaaaaaaataaaaaaacatagagaataatgaatgaattatgaagaagaatg 780
721 TACATTAAAAAAATAAAAAACATAGCAATATGAATTAATGAATTAATTAATGAGAA 762
781 aaaaacattgataaaaaataagaatgcaactaagaagaagaagaagaagaataatatacaa 840
781 AAAACATTGATCAAAATTAAGAAATGCAATATGAAGAAAGAAAGAAATTAATATCA 822
841 gctcaatagatccttctattcaaatatacaatagagaagaagcacataatataagc 900
841 GCTCAATATATCTTTTATTTAACAATAACAATTAACAAGACATTAATTAATAAC 882
901 gttttagaagaagaatgtagacactttaaaaaaaataatgaagaacatgaagaattactg 960
901 GTTTTGAAGAAAGATTAATGACACTTTAAAAAAATGAAGACATTAAGAAATTA 942
961 aagataatagaaattaaaa--atcccccacgycgaacttctggaatatacacaactact 1017
961 AAGATATGATTAATAATTAACAGATGCCGAAATCTCATCTAGGAGTAAACCAATCT 1002
1018 ctccctg---ataagaacaaaaaatcgaggaacacgaaagaagaataaagaatctgc 1074
1018 CTCCCTG---ATAAGAACAAAAATCGAGGAACACGAAAGAAATTAATAAGAAATTC 1062
1075 aaaaacttaaatatgaattatgataagttattactgacacttgaatataataactat 1134
1075 AAAACTATTAATTTTACATTTGATGATTTTATTTACTGATCCACTGTAATTAATTA 1122
1135 ttaagaagaataaataaataatgatalaagtgcaaaaggttga-----acaaag 1185
1135 TTAAGAGAAAAAATAATTAAGAAAGTGTGATGAACACCTTAATCAACAAGATCCTGA 1182
1186 gaatcaactgaaccacaaatgaatatccaaatgtagtacttataccttgcataaagat 1245
1186 GATCAATATCAATTTTACATTTGATGATTTTATTTACTGATCCACTGTAATTAATTA 1242
1246 attaaacatgctta-----aalgaaactaatcttcttggtggaattatataatcca 1296
1246 ATTCAATATTTCAATTTAGCTGACATTAATGATTAATAATTAATGATGATTAATTA 1302
1297 ttgattatatacaagaagaacaaagtaaaaaacatatacagatataatgaagaagaatcc 1356
1297 TTGATTATATACAAAGAAATTAATGAAGAAATTAATTAACAGATAATTAAGAAAGAA 1362
1357 atcaatgaataatgaagaagaataatlaaatagaagaagaagaagaatggaatctgataa 1416
1357 ATTAATTAACATTTTAATAAATAAATAATGATTTAGTAAGAAAGAAACATTAATTA 1422
1417 tcttaagaagaagatccttagcttcttaaatgatatatacaaaagaagaataatgaat 1476
```

```

1423 CAAAATTAATAATTAATCTTGAAGATTATGAAAGTCAAAAAAGATTATGAAGATTCTT 1482
1477 aatgaattatgatagcaaatccaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 1536
1483 GAAAAATTTTATGAAATGAAATTTTAAATTAATTTTGCACAAAGATGTCGATGAATATA 1542
1537 atgggtataaagatatcatataaaagtctgaagaacttcacacacataactcttgcatcc 1596
1543 TTCAGTGCAGATATACATATTAATGTTGAAACCAAAAAATATATATTAATTTTCTCC 1602
1597 tatgaataatcctaacaatactctgaagaacttcacaaagctcttaaatatctgaagat 1656
1603 TCATAATTAATCTGATATTAATTAATGTTCAAAAAATTAATAAGCTCTTCTATCTTGAAGAT 1662
1657 tatcttttaagagatatagtagtctgaagaagaataaataatataaataaataaataagc 1716
1663 TATTTCTTTAAGAAAGAAATTTCTGAAAAAGATTATTAATCATTTATTAATCTTGAAGAACT 1722
1717 aaaaatgaataatgagattgaaacatctagcttgaaataataaataaagaatgaagacagctc 1776
1723 GGCCTCGAAGCTGATATTAATAAATTTACGAGAAATTAAGAGTAGTGAACAAATTT 1782
1777 ttggaataaataatctactaagaacgaataataaaccagatgaataaattttagaagatct 1836
1783 TTTGAAAAAATAAATTTTAAAGACGTAACACATTCAGCAAAATGCTTCTTGAAGATCT 1839
1837 gaaatttgaagaatgacaagaatctgaagaagaatttgaagaagaagaatctgacgaataa 1896
1840 GATATTGTAATAATTAACAGACAAAGCTTTTATTAATTAATAAATAAGAGACTTTAGA 1899
1897 aagacacaaatctgaattttaaanaaattgaagaataaataataataataataataataata 1956
1900 AAGATGAGATTTATTTTAAATAATGCAACAACTAAAGATGATGATGATGATGATGATGAT 1959
1957 taacaacaagaataataagcaagaacattatctatctatctatctatctatctatctatct 2016
1960 TATAAATCAACAAATTAACAGCAACATATTAATTAATTAATTAATTAATTAATTAATTAAT 2019
2017 aattttaaagaatgcttgcctgaagaatgaataatgaagaagaagaagaagaagaagaaga 2076
2020 AATTTAAAGAAATTTTATTAACAAAGTAATAAGACATGTTAAAGAAAGACAGCTGCTTTA 2079
2077 aaaaacgaagaatgcaatcgaataatctgaaacacaaacgaagaagaagaataaagaagaca 2136
2080 TCAGATATTACACAACTTATGTTAGTGCAGCAAGGAAAC---AAGTGAAGATGGGGGTCTAC 2136
2137 gcaaatcaaaaacatctgaaacaaagaacagatctgctttagaagaagatctgatacaga 2196
2137 TCACACACACATTAATCCCAATCAGAGAAACAGAAAGAAACAGAAAGAAACAGTAGACAC 2196
2197 caagcaagaagaataaagaagaagaacacagtagcagtagcagtagcagtagcagtagcagtag 2256
2197 ACAACACAGGTAACTAATACATTTACCAACACACAAACATCACCACCAAAAGAGTAATAA 2256
2257 gcaacaagctccaacacacacacagcaacagtaataataaactgaataatgcttccaatca 2316
2257 GTTGTGTAATAATTTCAATAGACATTAAGAGATGACAAATTCACACAGCTTGCACAAATAAC 2316
2317 gattatcttgaagaataatctatctatctatctatctatctatctatctatctatctatct 2376
2317 GTTTATCTAATAAGAAATTAATGAAATTTTAACTAAATTCATTAATTAATTAATTAATTAAT 2376
2377 ttggtttcacctcaacttgaagaagaagaatgattaaacaataaataaataaagaagag 2436
2377 TTAGTATCAAACTCTAGTATGACCAAAAAATTAATTAAGAGTATATATTAATCTTCCAGAA 2436
2437 gaagaagaacaataatgaatctatctatctatctatctatctatctatctatctatctatct 2496
2437 GAAAAAATGAAATTAATAATCATGTGATTCATTAATTAATTAATTAATTAATTAATTAATTA 2496
2497 ataactgtatctatctatctatctatctatctatctatctatctatctatctatctatct 2556
111111 111111 111111 111111 111111 111111 111111 111111 111111 111111

```

3559 aaattctccagaaggaacagat----- 3582
3634 AATTTCCTCCAGACGAAAGTTTACACAGTTGTACTCCACCTCAACCAATGTAAT 3693
3583 ---gttcaacagcttglaaglaagtgatccgcacacaltgaacaaagtcacaag 3639
3694 CCATCTCCATTTATCTGTAAAGGTAAGTGTAGTTCAGATCCCAAAAGAGAACACA 3753
3640 aaacagacatcaactcattgtagagagctcaacacatacaacacacaaatgtc 3699
3754 AATCCAACTTCAGGCTCTTATTATTAACAACTTAACAACTTAACAACTTAAT 3813
3700 gattgtagaagtagagcattacatcaatcattatcttggaagacccaagaagttat 3759
3814 GACGAGAAAGATGATTCCTTAGTGTATTAATCCATTTTGGAGAAATCCGAGATTAATGAC 3873
3760 gatgatttagaagaagtaglaacagaagaagcagtaactcttcglaattgatacata 3819
3874 GAATATTTAGATCAAGTACTGAGAAAGCAATA---TCTGTCACAATGATTAATATC 3930
3820 cttctcaaaatggaagaatgaggtttatattataaacccttagcaggtgttat 3879
3931 CCTCTAGGATTTGAAATGAAATGATGATTTATTAATTAACCTTAGCTGAGATATAT 3990
3880 aagaagttiaaaaaaacaattagaanaatacgttatgacattlaatgttaagat 3939
3991 AGAAGCTTTAAAAAACAAATTGAAAAAACATTTATTCATTTTAATTTGAACGAT 4050
3940 atttlaaatccagattlaataaagctgaaatltcaaaatgttttagatcagattla 3999
4051 AACTTAATTCACGCTTTAAGAAACGAATAATTTCTTGATGTATTAAGATGTGATTTA 4110
4000 attcctaagaagattcaacacaglaattatgtgtcgaagatccataaattctc 4059
4111 ATGCATTTTAAACATATATCCCAATGAAATGAAATGAAATGAAATGAAATGAAATG 4170
4060 aataaagaanaagagataaattcctaagcagttataatataatgaagtaagat 4119
4171 AATTGGAACAAAAAACACACTTTTAAAAAGTTTAAATATTAAGATACAGTAGAA 4230
4120 acgagataaattttgcaaatgagtgctcttgatataataatataatccgaataat 4179
4231 AATGATATTAATTTGACAGGAAAGTATTAAGTTATTAAGTTATTAAGTTATTAAGTTAT 4290
4180 aaatcagattgattcattcaataaataatataca----- 4214
4291 AAGGATGATTTAGAAATCAATTAATAAGTTATCAAAAGAAAGAGTTCCCATCATCA 4350
4215 -----cgacaacaaggtgaanaatgaagtaac 4242
4351 CCACCAACAACCTCTCGTACCAGCAAAAAACAGCAACAAAAAGAAAGTAAGTTC 4410
4243 ctctccctttcaacaatltgagacctatataaacaagttatgaataatgtatta 4302
4411 CTTCCATTTTAAACAACATTTGACCTTATTAACATTAATTAATTAATTAATTAATTAAT 4470
4303 ttgttaattcattagaagaagaagttcttaaatatatacatatagaagaatcaacgtlagaa 4362
4471 TACTTAATTAACCTTAAGCAAAAGTTAAGATGTAATGTTGAAAAAGATGAACACAT 4530
4363 gttaaataaagaacttaataacttaaaaaaacatcaagacaaatgtgcagaatttaaa 4422
4531 GTTAAATTAACCTTAAGTGTGATTTAAAAAGCAATGATGACAAATAGATVCTTTTAAA 4590
4423 aaaaataacaattcgttggaattgctgattatcaacagattataacataacta 4482
4591 AACACATAGACTTCGACACAAATTAATAATTAATGATATACGAAAAAGATATG 4650
4483 ttgaacaagttcccttagtaacaggtatgttttgaanaatcttgtaaaacggttattct 4542
4651 CTTGGCAAAATTTACTAGTACAGGATTAG---TTCAAAATTTTCTTAATACAAATATATCA 4707

4543 aattlaattgaaacttgcaggtaglaltlaaacattcacacaaccaatgcgtaaaa 4602
4708 AATTTATTAACGAATAATTCGAAGATATGTTAAACATTTTCAACACCAATGCGTAAAA 4767
4603 aaacaatgtccacaanaattctgtgatttccaacatttagatgaagaagaatgtaa 4662
4768 AAACAATGTCCAGAAAATCTGTGATGTTTCAGACATTTAGATGAAGACGAATGTAAA 4827
4663 tgtttatlaattacaacaagaagtgatataatgtgttgaanaatccaaatcctactgt 4722
4828 TGTTTTAAATTAACAACGAAGATGATTAATGTTGAAAAATCCAAATVCTTACTGT 4887
4723 aacgaataatgtgtagtgaatgcagatgccaaatgtlaecgaagaagattagtagc 4782
4888 AACGAAATATATGCTGATGTGATGATGATGATGATGATGATGATGATGATGATGATG 4947
4783 aacggaagaanaatcacatgtgactaactaactgattatccactttcagatgt 4842
4948 AACGAAAGAAATCAACATGATGATGATGATGATGATGATGATGATGATGATGATG 5007
4843 atttctgcagctcccttaactccttagaataatcattcattataactcagttata 4902
5008 ATTTCTGCAAGTTCCTTAACCTTTAGGAATATCATTTTATTAATATCATGATTAATA 5067
4903 ttatacagttcattaa 4920
5068 TTATACAGTTTCATTTTA 5085

RESULT 11
AF062348
LOCUS
DEFINITION
Plasmodium falciparum strain HNI merozoite surface protein 1
precursor (msp1) gene, complete cds.
ACCESSION
AF062348
VERSION
AF062348.1 GI:3859842
KEYWORDS
SOURCE
malaria parasite P. falciparum.
ORGANISM
Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE
1 (bases 1 to 5243)
Jiang,G., Liu,R., Daubenberger,C.A. and Pluschke,G.
Sequence analysis of the MSP 1 gene of Plasmodium falciparum from
Hainan, China
JOURNAL
Chung Kuo Chi Sheng Chung Hsueh Yu Chi Sheng Chung Ping Tsa Chih 17
(5), 294-297 (1999)
REFERENCE
2 (bases 1 to 5243)
Jiang,G., Liu,R.Z., Daubenberger,C.A. and Pluschke,G.
Direct Submission
JOURNAL
Submitted (30-APR-1998) Molecular Immunology, Swiss Tropical
Institute, Socinstrasse 57, Basel CH-4002, Switzerland
LOCATION/Qualifiers
1. 5243
/organism="Plasmodium falciparum"
/strain="HNI"
/db_xref="taxon:5833"
/chromosome="9"
/country="China"
/gene="msp1"
/product="merozoite surface protein 1 precursor"
/gene="msp1"
/gene="msp1"
/note="binds spectrin"
/codon_start=1
/product="merozoite surface protein 1 precursor"
/protein_id="AAC72884.1"
/db_xref="GI:3859843"
/translation="MKIIFFLCSPLFIINTOCVTHESYOELVKRLKLEDALEDAVLNGYS
LFOKRMVINECTSGAVTSTPGSGSVTSVSGSGSVASVSGSGSGSVASVSGSGS
RTINPSDNSDSDAKSYADLKIRVQNYLFTIKELVPELFDLTNHLTLCDNIHGKY

1657 taattcttaagaagaatagtagtgaagaaagaatlaaaatlatlaaataaattlaaagc 1716
1728 tttcttttaagaagaatgattctgaaagaaagatttaacattatattatcttggaaact 1787
1717 aaataagaagaatgagatggaacattgltgaaatalttaaaagaagaagacagctt 1776-
1788 ggcctcgaaagctgattttaaataaataaagaaatgaaagagagtgagaaacaaat 1847
1777 ttgaaaaaaataattactaaagacgaataaaccagatgaaataaattlaagaatctt 1836
1848 ctgagaaaaa---tttttaagacctaacacattgacaaatgcttcttaagatatt 1904
1837 gacattgaaagtagaagttcaaaaagtttattcaatgaacaaatgacgaatga 1896
1905 gatatttgtaaatattacacattacaaagcttttttttttttttttttttttttttttt 1964
1897 aagactcaattgattttaaataatgagaatlaaaacataatacatatgltcccaattct 1956
1965 aagatagaatt 2024
1957 tacaacaaagaataaagaacaaaccttatatttaattgltgaaagaaagaaatgtat 2016
2025 ttatataacacaaataaataaaccacacacatttttttttttttttttttttttttttttt 2084
2017 aaattaaagttgcatgcttaagtagaatacatgataatgaagaataaataaatacata 2076
2085 aaatt 2144
2077 aaacagaagtgcaatcgagataatcgagacacacacacacacacacacacacacacac 2136
2145 tcaagattatcac 2201
2137 gcaactacaaac 2196
2202 tccac 2261
2197 caagcagaagaacaaataaagaacacacacacacacacacacacacacacacacacacac 2256
2262 ac 2321
2257 gac 2316
2322 gttgtgtaaaatttcaatgaacatgaatgaatgaatgaatgaatgaatgaatgaatgaat 2381
2317 gattatctgaaataatlaagaatttlaaactacataatgltcaacaaatatt 2376
2382 gtt 2441
2377 ttggtttcac 2436
2442 ttatgattcaaacctgattgagacacacacacacacacacacacacacacacacacacac 2501
2437 gaagaagaacaaatlaagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 2496
2502 gaaaaaaatgaattttaaattcatgctgattgattgattgattgattgattgattgattgatt 2561
2497 ataactgtaagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag 2556
2562 atactgctgattgattgattgattgattgattgattgattgattgattgattgattgattg 2621
2557 gaaatttgaagaagaatggtgattgattgattgattgattgattgattgattgattgattg 2616
2622 gatt 2681
2617 aaaaaattatlagagaagcgaaaaaa-----gtatccacacacacacacacacacacac 2670
2682 aaaaaattatlagagaagcgaaaaaa-----gtatccacacacacacacacacacacac 2741
2671 agtcaatgaatgac 2717
2742 atatacaac 2801

2718 -----agta 2721
2802 aatgacatctctac 2861
2722 agtgcac 2781
2862 gaagaagaac 2921
2782 aacatattgagtcctgggaaaaaataatatacccaagatta---ataggtcaaaa 2838
2922 actgcttttaatttggaaatataaacttaagtagccttaacacacacacacacacacac 2981
2839 agtgcac 2898
2982 gatttgcac 3041
2899 ttac 2958
3042 atcaaaac 3098
2959 aggaagaac 3018
3099 aaaaaac 3158
3019 ttata 3078
3159 ttatataataataataataataataataataataataataataataataataataataata 3218
3079 aaata 3138
3219 caaagac 3278
3139 aattcaacttaataataataataataataataataataataataataataataataataata 3198
3279 aattcaacttaataataataataataataataataataataataataataataataataata 3338
3199 aagaagaac 3258
3339 aagaagaac 3398
3259 catlaaagaac 3318
3399 cattatataataataataataataataataataataataataataataataataataataata 3458
3319 gaagaac 3378
3459 gaagtttcaatttcaac 3518
3379 aatttagaagaac 3438
3519 aaaaatgac 3578
3439 tcaagtgcac 3498
3579 tcaagtgac 3638
3499 acagtgac 3558
3639 acagtgac 3698
3559 aatttccac 3582
3699 aatttttccac 3758
3583 ---gtgcac 3639
3759 ccatttccac 3818
3640 aac 3699
3819 ataccac 3878
3700 gatgatgaagtagatgac 3759

```

Db 3879 GAGCAAGAAGATGATTCCTAGTGTATTAATCCATTTTGGACAATCCGAGATTAATGAC 3938
Qy 3760 gatgatttaggacaagtagtaacagagaagcaagtaactcttcgcgaattgatacata 3819
Db 3939 GAAATATTTAGATCAAGTAGAATCACTGAGAAAGCAATA---TCTGTCAACATGATTAATATC 3995
Qy 3820 ctcttcaaaattgaataatgaatgaaggttttaatttaaacctttagaagtgattat 3879
Db 3996 CTCTCGAGATTTCAGAAATGATATGATTTATATTTAAACCTTAGCTGAGATATAT 4055
Qy 3880 agaagtttaaaaaaacaatagaaaaaatacgtltatgaacttaattgaatgaagat 3939
Db 4056 AGAAGCTTAAAAAACAATTTGAATAAACATTATTACATTAAATTAAATTGAACGAT 4115
Qy 3940 attttaaattcagatttaataacgttgaaaattccaataatgctttagaactcagatt 3999
Db 4116 ATCTTAATTTACGCTCTTTAAGAAACGAAATATTTCTTACATGTATTTAGAACTGATTTA 4175
Qy 4000 attcatalaagaatttaacataagtaattatgctgtcaagaatccataaatttctt 4059
Db 4176 ATGCATTTTAACATATATCTCAATGATACATTATTGAAGATTTCAATTAAATTATG 4235
Qy 4060 aataaagaanaaagaataaattccttaagcagttataatataatgaattcaatagat 4119
Db 4236 AATTGACACAAAAAACAACACTTTTAAAAAGTTACAAATATATATAAAGAACAGTAGAA 4295
Qy 4120 acgagataaatttgcgaatgctgtcgtgatatataaataatataatccgaataat 4179
Db 4296 AATGATATTAATAATTTGCACGAGAGATATAGTTATTTATGAAAAAGGTTTACGAAATAT 4355
Qy 4180 aatcagatttagatccaatttaaaatatataa----- 4214
Db 4356 AAGCATGATTTTAAATCAATTTAAAAAGTTATGCAAGACAAAAAGAGAGTCCATCATCA 4415
Qy 4215 -----cgaaaaaagaagtgaaaaatgaatc 4242
Db 4416 CCACCAACAACACCTCCGTCACCAGCAAAAAACAGACGAAACAAAAAGAAAGTAAGTTTC 4475
Qy 4243 ctctcccttttaacaatattgagacttatataaacaagttatgataaattgattta 4302
Db 4476 CTTCCTATTTTAAACAACATTTAGACCTTTATACAAATTAATTAATTAATAATTGACGAT 4355
Qy 4303 ttgttaattcaattagaagaagaagttctaattatacatatagagaatccaacgttagaa 4362
Db 4536 TACTTATTTACTTTAAAGCAAAAGATTAAAGATTGTAATTTGAAAAAGATGAGACAT 4395
Qy 4363 gtttaaaataaagaacttaattacttaaaaaacaattcaagaacaatttgcagatttaaa 4422
Db 4596 GTTAAATATTAACATACTTACTGATTAAAGCAATTGATGACAAATAATGATCTTTTAAA 4655
Qy 4423 aaaaaatacaatttgcgttggaattgctgatttatacaacagatatataaccataactta 4482
Db 4656 AATACCTAACGACTTGAAGCAATTTAAATAATTTGATTAATGATATGAGAAAGATATG 4715
Qy 4483 ttgacaagaattctccttgtagaagtagtcttggaaaactctgtaaacacgtttatct 4542
Db 4716 CTTCGCAAAATTTACTTGTAGTACAGATTAG---TTCAAAATTTTCTTATATACATAATATCA 4772
Qy 4543 aatttacttgatggaaacttgcgaaggtatgttaaacattcacacaacacacgttaaaa 4602
Db 4773 AATTAATTGAAGAAATAATCCAAAGATATGTTAAACATTTCCACACACCAACATCGTAAAA 4632
Qy 4603 aaacaatgtccacaataatcttgagatgttcagacatttagatgaagaagaagaatga 4662
Db 4833 AAACCAATGTCCAGAAATTTCTGATGATTTTCAGACATTTAGATGAAAGAGAAATGTAAA 4692
Qy 4663 tctttaaattacaaacaagaaggtagataatgtgtgaataatccaactcactactgt 4722
Db 4893 TGTTTATTAATAATTAACAACAAGAGGTGATAAATGTGTGAAAAATCCAAATCTCTACTTGT 4952
Qy 4723 aaagaaataatgtgtgagatgtgacagatgccaatgttaaccaagaagaattcagtagc 4782

```

```

Db 4953 AACGAATAATAGTGATGTGATGCAAGATGCCAANTGTACCAAGAACAATTCAGGTAGC 5012
Qy 4783 aacggaagaataatcaatgtgaaatgactaaacctgattctatacacttgcagtgat 4842
Db 5013 AACGGAAGAAATATCACATGTGATGATCTAAACCGATTTCTTATTCACCTTTGAGTGT 5072
Qy 4843 atttgcagatcccttaactctcttagaataatcattcttaataactatgataata 4902
Db 5073 ATTTTCTGCAGTTCCCTTAATCTTTAGCAATATCAATTTCTTATTAATCACTGTTAATA 5132
Qy 4903 ttatacagttcaattaa 4920
Db 5133 TTATACAGTTTCATTTAA 5150

RESULT 12
PEP190G1
LOCUS PEP190G1 5392 bp DNA INV 22-JAN-1999
DEFINITION Plasmodium falciparum p190 gene for surface antigen precursor.
ACCESSION X05624
VERSION X05624.1 GI:4186080
KEYWORDS antigen; glycoprotein; p190 gene; signal peptide; surface antigen.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 5392)
Tanabe, K., Mackay, M., Goman, M. and Scaife, J.G.
Allicic dimorphism in a surface antigen gene of the malaria
parasite Plasmodium falciparum
J. Mol. Biol. 195 (2), 273-287 (1987)
88011243
REFERENCE 2 (bases 1 to 5392)
Tanabe, K.
Direct Submission
Submitted (21-JAN-1999) K. Tanabe, Osaka Institute of Technology,
Biology, 5-16-1 Ohmiya, Asahi-Ku, Osaka, 535-8585, JAPAN
REMARK Sequence revised by author
COMMENT On Jan 26, 1999 this sequence version replaced gi:9926.
FEATURES
location/Qualifiers
1..5392
/organism="Plasmodium falciparum"
/strain="MAD20"
/db_xref="taxon:5833"
/clone_lib="(lambda)NM1149 genomic"
136..192
136..5241
sig_peptide
CDS

```



```
ENEVDYIKPLAGVYRSILKKOIEKNITFENLNDILNSRLKRRKRYELDLVLESDLMQ
FKHISNEYIIEDSEFKLNSDEQNTPLSKYKIKSEVENDIKFAOEGISYKYLAKY
KDDSEIKVEIKKEKEKPEPPTPPSPAKTDEQKESKFLPLNIETLYNNLVNK
IDDYILINKAKINDCNVEDEAHVHKITRSLDKAIDDKRIDFKRNDPEAIKKLIND
TKKMLGKLLSTGLQVONFPNTLSKLIBEKPDMINISDHQVKKQCPENSCFRLHD
EREKCLINRKOEBCDKEVNPNCNENNGCGDADATCTEEDSSSRKKTICECTKP
DSYPLFDGIFC5SSNFI6ISFLIIMLILYSEI"
193. .5238
mat_peptide
/product="surface antigen (AA 1-1682)"
463. .465
misc_feature
/feature="pot.glycosylation site"
850. .852
misc_feature
/feature="pot.glycosylation site"
1543. .1545
misc_feature
/feature="pot.glycosylation site"
1741. .1743
misc_feature
/feature="pot.glycosylation site"
1954. .1956
misc_feature
/feature="pot.glycosylation site"
2539. .2541
misc_feature
/feature="pot.glycosylation site"
2575. .2577
misc_feature
/feature="pot.glycosylation site"
2830. .2832
misc_feature
/feature="pot.glycosylation site"
2880. .2882
misc_feature
/feature="pot.glycosylation site"
2890. .2892
misc_feature
/feature="pot.glycosylation site"
3028. .3030
misc_feature
/feature="pot.glycosylation site"
3106. .3108
misc_feature
/feature="pot.glycosylation site"
3400. .3402
misc_feature
/feature="pot.glycosylation site"
3733. .3735
misc_feature
/feature="pot.glycosylation site"
4897. .4899
misc_feature
/feature="pot.glycosylation site"
5032. .5034
misc_feature
/feature="pot.glycosylation site"
BASE COUNT 2376 a 685 c 702 g 1629 t
ORIGIN
Query Match 55.58: Score 2732; DB 96: Length 5392;
Best Local Similarity 73.98: Pred. No. 2.3e-245;
Matches 3795; Conservative 0; Mismatches 1095; Indels 246; Gaps 16;
1 atgaagatcatactcttctatgctcatctcttcttattataataacacatgtgta 60
Db 136 ATGAAGATCATATCTCTTTTATGTTATCTTTTATTTATTAATACACATGTGTA 195
Qy 61 acacatgaagatlatcaagaactgtcaaaaactagaagctttagaagatgcagtatg 120
Db 196 ACACATGAAGAAATTATCAACAAGCTTGTCAAAAACCTGAGAGCTTTAGACAGTGAATTG 255
Qy 121 aaagttatagttatttcaaaaagaaaatggtatataatgaagaagcaagtgtaaca 180
Db 256 AAGAGTTATAGTTTATTTTCAAAAAGCAAAAATGATTAATGAAGAACAACTGGAGACA 315
Qy 181 gctgttaacaactagtcacacctgttcaaaaggttcagttcagttcaggtgctcagtg 240
Db 316 GCTGTTAACAACAGTACACTGTTCAAGTGTGTTCAAGTACTTCAAGTGTG----- 365
Qy 241 tcaagttcctcaggtgcagttgctcaggtgctcagttgctcaggtgcagttcagtt 300
Db 366 -----TTTCAGTTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 417
Qy 301 gcttcaggtgttcaggtatctcaagcgtacacatccttcagatattcaagtgatca 360
Db 418 GCTTCAGGTGTTCAGGTATCTCAAGACTTCAAAATCTTCAAGATTAATTCAGATGATTA 477
```

```
Qy 361 gatgtcaaatcttaagctgtatttaaaacacagagtaacgaatattgttaacatcaaa 420
Db 478 AATAGTAAACCTTACCGCTGATTTAAACATAGAGTTCAAAATATTACTGTTCACATTA 537
Qy 421 gaactcaaatatcccaactccttgatttaactaatatgttaactctgtgtgataat 480
Db 538 GAACCTAAATATCCCAACCTTGTGATTAACCAATCAATATGTGACTTATCTAAATA 597
Qy 481 atcattagtttcaaatatttaattgatlgaatagaagaatattgaattatataaa 540
Db 598 GTTGATGATTTTCAAAATATTAATTTGATGATGAAGAAATTAATTAATTAATTA 657
Qy 541 ttaactttatttattgattatttaagaagcaaaatgaatgatlgtgtcaatgattat 600
Db 658 TTAACTTTTATTATGATTTATTATTAAGACCAATTAATGATGATGATGATGATGAT 717
Qy 601 tgtcaaatccttcaactcaacttaaaatcgtgcaatgaattagacgtacttaaaaa 660
Db 718 TGTCAATATACCTTTCAATCTTAATAATTCGTGCAATGAATTAACGACTTAATAA 777
Qy 661 gtgttcgatatagaagaaccattagacataatagaagaatgtagaagaatgagaat 720
Db 778 GTGTTGGAATATGAAGAACCATTAAGCAATTAATTAAGATTAATGAAGAAATGA 837
Qy 721 tacattaaaaaaataaaaaacacacacacacacacacacacacacacacacacac 780
Db 838 TACATTTAAAAAATATTAACACACATGACAAATTAATTAATTAATTAATTAATTA 897
Qy 781 aaaaacttgaataaaaaataagaatgcaactaagaagaagaagaagaagaagaaga 840
Db 898 AAACAATGATGATCAAAATTAAGATGACAGATTAATGAAGAAAGAAAGAAAGAA 957
Qy 841 gttcaatgatctctctctatctatcaataaacaattagaagaagcacatataatga 900
Db 958 GTCGAATATATATCTTTTATTTTATTAATTAATTAATTAATTAATTAATTAATTA 1017
Qy 901 gtttgaagaacgtaatgacactttaaataaaatgaagaacatgaagaatattgatt 960
Db 1018 GTTTTGAAGAAAGCTATTGACACTTTAAAAAATGAAGAAACCTTAAGAAATTA 1077
Qy 961 aagataaatgaataaaa--atccccaacgcgcacacttctgaaatatacacaat 1017
Db 1078 GATATGATTAATTAATTAACAGATGCGCAAAACCCCACTACTGAGATTAACCAAT 1137
Qy 1018 ctctctg---ataagaacaaaaaatcgaaggaacacgaagaagaagaagaatctgc 1074
Db 1138 CTCCCTGAGAAATTAAGAAAGATGATGATGATGATGATGATGATGATGATGAT 1197
Qy 1075 aaaaacttaaatattatagatgattatctactgacacttgaattagaatactat 1134
Db 1198 AAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1257
Qy 1135 ttaagaagaaaaaataaaaatataatgaatgaatgaatgaatgaatgaatgaatga 1185
Db 1258 TTAAGAGAAATAAATAAAGATGATGATGATGATGATGATGATGATGATGATGAT 1317
Qy 1186 gaatacactgaaccccaatgaatataccaatgaatgaatgaatgaatgaatgaat 1245
Db 1318 GTTCAATATCAAAAGCTTCTTATCAAAATGATGATGATGATGATGATGATGAT 1377
Qy 1246 attaacaatgcttta-----aatgaactaatctcttggtgatttaataatca 1296
Db 1378 ATTCAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1437
Qy 1297 ttgattatatacaagaacaaaglaaaacatataactgaatgaatgaatgaatgaat 1356
Db 1438 GATACATAAAGAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1497
Qy 1357 ataaatgaatataagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1416
Db 1498 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1557
Qy 1417 tcttaagaagaagatcactaagcttcttaaatgatalatacaaaagaatatgaat 1476
```

Db	1558	CAAAATTAATAAAATTA	CTTGAGAGTTATGAAAGTCA	AAAGTC	AAAAAGGATTATGAAATTA	CTT	1617
Oy	1477	aatgaattatctgacgaattcaatcaataatataatgatttactcaatttcgaaaatg					1536
Db	1618	GAAAAATTTTATGAAATGAAATTTAATATATATTTT	GACAAAGATGTGTGAT	TAATA			1677
Oy	1537	atgggtlaaagatctatcatataagtttggaacctacacccataactttgtatcc					1596
Db	1678	TTTCAGTGCAGATTTACATTTAATGTTGAAAAACAAGATATATAATTAATTTTTCATCC					1737
Oy	1597	tatgaaatcttaacgaatactctggaaggttacaaaagcccttaaatatetggaagt					1656
Db	1738	TCTATATTTCTGATATTAATGTTCAAAATTTAAAAAGGCTTTTCATATCTGAAAT					1797
Oy	1657	tattcttlaaggaatcagtcagttctgaaaaagaatlaaaatctataaaatttaaaagc					1716
Db	1798	TATTTCTTTAAGAAAGGAATTTCTGAAAAAGATTTTAATCATTATTAATCTTGAAACT					1857
Oy	1717	aaatatagaatactgatctggaacatagttgaaatalataaaagaatgaagacagct					1776
Db	1858	GGCTCGAAGCTGATATATAAAAAATTAACGAGAAATTAAGAGTAGTGAACCAAAAT					1917
Oy	1777	tttaaaaaaaatctactaagaagcgaataataacccgatgaaaaaatttggaagtct					1836
Db	1918	CTAGAAAAA---TTTAAAGACTACACATTCGCAAAATGCTTCCTTAATAATATCT					1974
Oy	1837	gacatctgtaaaagtcacagttccaagaatttctaaatgaacaaattgaacgaattaaa					1896
Db	1975	GATATTTGTAATAATTNCAAGTACAAAAAGTTTATTTATTAATAAAAAATAGSAGCTTAAGA					2034
Oy	1897	aagctcaattgattcttaaaaaatctgaatatlaaaacataatacatgcttccaattct					1956
Db	2035	AAGATAGATTTATTTTAAAAAATGCAACAATTAAGATAGTATTCATGTACCAATATTT					2094
Oy	1997	tacaaacgaagaaatgaagacagaccttatctaatgtgtgtgaaaaaagaattgat					2016
Db	2095	TATTAACCCCAAAATTAACCAAGAACCATATTAATTAATGTTTAAAAAACAATCAAT					2154
Oy	2017	aaatttaaagtggttcagcttaagtagaatcatctgataatggaagaaaaaaacaata					2076
Db	2155	AAATTTAAAGATTTATACCAAAAGTAAAGACATTTTAAAGAAAGAACACAGCTGTCTTA					2214
Oy	2077	-----aaacagaagctcaatcgcgatctcgaaaccaaccaacggaaggaataaca					2130
Db	2215	TCAAGTATTTACACAACTTTTAGTTGCAGCAGGAAAGAACATGAAGATGGGGGTACATCC					2274
Oy	2131	ggacaagcaactataaaaacttggaacaaagccggatctgccttggaaggagatctcgta					2190
Db	2275	ACACACACATTATTCATATAGAGAGAAACGGAAGTAAAGAAACAGAACTAACAA					2334
Oy	2191	caagcaagaacacagaacaaaacgaacgaacacccagtcacagtcacagtcacagaa					2250
Db	2335	GAAACAGTAGACACACACAAACAGGTTACATTAACCTTCCACCAAGAAAGAAATCAGCA					2394
Oy	2251	gcaaaagcacaaagtcaccaacacccaagcaccgaataataataaaactgaaat-----					2304
Db	2395	CCAAAGAGATGAAGTGTGTTGAAATTTCAATAGAACATTAAGATAGACATTCACAA					2454
Oy	2305	---gttccaaatagatctatcttgaaaaatataatgatttttaataactcatata					2361
Db	2455	GCCTTGACAAAACAGTTTATCTTAAGAAATTAAGATGATTTTAACTAATATCATATTA					2514
Oy	2362	tttcaacaattatcttggtttcgaacctcaactatgaagaagaagaatcttaacaatat					2421
Db	2515	TGTATTAATAATTTATTTATTAATTAAGTCAACTGATATGACCAAAAAATTTATAGAGTATAT					2574
Oy	2422	aaattcaaaaggggaagaagaacaaatgaagtctgacccaattgaactatggtt					2481
Db	2575	AATCTTACTCCAGAGAGAAAGAAAGATTAATTAATCTATGATCCATTAAGATTTATATTT					2634
Oy	2482	ataatacaaatatacctctgataatgctatctgattgtagcttgaagaatgattta					2541

Db	2635	AAATTTCAAAATTAACATACCTGCTGATGTATTCTTTATTGATAGATGATGACATGATTTA	2634
Qy	2542	tcacacacatttacttgcgaatttctatgcgaagaagatggtttgttaattctataaaccttaag	2601
Db	2695	CAACATCTCTTTTGTGAATTAATATCAAAAAGGAATGATTTATTATTAATACATTAACATAAA	2754
Qy	2602	gataatgcgaaaatttaaatttattatagggaagcgaaaaa-----gtatccacactc	2655
Db	2755	GAGGAAATTCATCAAAAATTTATTAAGAGACCAAAAACAAATTAACGTGAACATCATCT	2814
Qy	2716	ga-----	2717
Db	2875	CAAAACCAACAAATCAATTCGATCCTTACCCAATACCCAAATATGCTGTATCATCACT	2934
Qy	2718	-----agtaagtcgaatgatgatatacctacattccaaatttgaataatgt	2766
Db	2935	GGTCTGCTGTAGTTGGAAGAAAGCATGATCCCTTAACAGTATGTCTATTATAGTAACGAT	2994
Qy	2767	ttaaattattttgcgaacatcttgagctctgcgaagaacacaaatatataccagaagata	2826
Db	2995	TTGAAAGSTATTGTTAGTCTCTTAATCTTGGAAATTAACCTTAAGCTACTATTCATTA	3054
Qy	2827	---ataggtccaaaaagtagtgcgaaccttctatgcgaagaagatactaaagatgagtataca	2883
Db	3055	ACCATTTCTACACAGATGGAATAATTTTATGAGATATTTTAAAAATTAATGATAC	3114
Qy	2884	tttctataatgaatccttccaaattctgtataaactcaaaagctgcgatattatatactatcg	2943
Db	3115	TATTTTATGATGATATCAACAACTTCGTAAATCTTAATTCAAAAGTAAATTAACAGTTTG	3174
Qy	2944	aatgatgcatacaaaaggaagaattatgcgaagaatattatataattttaaanaaacctta	3003
Db	3175	A---CCGAACACAAAATAATGCACTTTAATGATGAATTAATAAATTAAGATACCTTTA	3221
Qy	3004	cagttatcatcttgctctataataataataataataatgcgaagaattatcttgatataa	3063
Db	3232	CAGTTATCATTTGATTTAG	3291
Qy	3064	aagaaacacgtctgttaaatataaatgcgaatttaaanaacttacttctataaagaacaa	3123
Db	3292	AAAAAAGAACTTGGCCAGACAAAATGCAAAATTAATAAACTTACTTTATTAAAGAACAA	3351
Qy	3124	ttgatatcaaaattggaatttaacttaataaccgaagctgattcttcaaaactttctgt	3183
Db	3352	TTAGATATCAAAATTTGAATTTCACTTAATATACCCATATAATGTAATTCACAAACTTTCGTT	3411
Qy	3184	ttctcttaacaaaaaagaagcctgcgaatagcgcgaacccgcgaaacacatgcgaagaaca	3243
Db	3412	TTCTTTTAACAAAAAAGAAAGCTGGAATTAAGCAATGCAAAACACATTTAGAAACACA	3471
Qy	3244	aaaatbatatgcgaacatctaaagagcttgttaaalattataatgcgtgcgaatcctca	3303
Db	3472	AAATATATATGCAAACTTTATAAGAGCTCTGTTAAATTTATTAATATGTCATCATCTCCA	3531
Qy	3304	ttaaaaacttgaagtcgaagaacaaatccaaacgcgaagataatctatgcgaattgcgaaac	3363
Db	3532	TTTAAAACTTTTAAGTGAAGTATCAATTTCAAAACGGAAGATTAATTTATGCAATTTTAAAAAA	3591
Qy	3364	cttaagatatttaagtaaatgcgaagaagaaattaaagatataatttaatttgcgaagaaga	3423
Db	3592	TTTGAAGCATTAAGTTAAATTAATGATGGAACCTCAATGATTAATTTTCAATTTAGGAAGCAA	3621
Qy	3424	aaatatacatcattatacaagtcgatctacatcaltttaatgcgcgaattaaagaaatlaa	3483
Db	3652	AAATTAATCTTTCTTAATCAAGTGAATTAACATCAATTTATTAACGATTTAAAGCAAGTATA	3711
Qy	3484	aaaaataaaaaattatcacaggttaactctccaagtcgaataataaagagatgattatacaatgc	3543
Db	3712	AAAAATTAATAATTTATACAGTAAATTTCTCAAGTGAATAAATTAATTAACAAAGTAAACCAACT	3771

```

OY 3544 ttgaatcttaaaaaaatctccccaagaagaacagat----- 3582
DB 3772 TTTAAATCTTACGAAATTTTTCCTCCAGAACAAAAGTTACACAGCTTTACCTCCACT 3831
OY 3583 -----gtgcacacagttgtaagtgaagtgagtcgcacacatagaa 3624
DB 3832 CAACGAGATGTAACCTCCATCTCCATTTATCTGTAGGGTAAAGGTGTTAGTTCAGAGATCCACA 3891
OY 3625 caaagtcaaccaagaagaaccgacatcaactcaltgtagagcagagtcatacacaataaca 3684
DB 3892 AAAGAGAAACCAATACCACTTCAGGCTCTTTATTAAAGAAATTACAAAGATGAGA 3951
OY 3685 acatcaaaaatgctgataagaatgaagtgacgtatcatctgcttatctttttgagaa 3744
DB 3952 CAATTACAAATTTATGACCAAGAGATGATTCCTTTAGTTGTTATTCCTCAATTTTGGAGAA 4011
OY 3745 tccgaagaagatcatgatctttagacacagtagtaacagagaagcaatlaactcctcc 3804
DB 4012 TCCGAAGATTAATGACGAATATTAGATCAAGTAGTAACCTGGAAGCATA---TCTGTC 4068
OY 3805 gtaattgataacatactctctaataatgaaatgaaatgaggtttatattataaact 3864
DB 4069 ACAATGGATTAATCTCTCAGGATTTGAAAATGAAATGATGATGATTAATTTAAACCT 4128
OY 3865 tttagcagtggttataagaagtttaaaaaaagaatagaataatgaatgaatgaat 3924
DB 4129 TTAGCTGGAGTTATTAAGAGCTTTAAACAAACAAATTTGAAAACATTAATTAATTAAT 4188
OY 3925 gttaatgttaagaatatttaaatcagatcttaataaagtcgaataatccaataatgtc 3984
DB 4189 TTAATTTGAACGATATCTTTAAATTCACCTTTAGAACAAAGAAATATTCTTAGATGTA 4248
OY 3985 tttagatcagatttaattcatalaagaatttaacacagtaattatgttgcagaagt 4044
DB 4249 TTAGATCTGATTTAATGCAATTTAAACATATATCTCAAAATGAATCAATTAATTTAGAGAT 4308
OY 4045 ccatlaaatcttctaataagaagaagaatgaatgaatcttaagcagttatattat 4104
DB 4309 TCATTTAAATTTTGAATTCAGAACAAAACACACTTTTAAAGTTACAAATTAATA 4368
OY 4105 aaggaatcatalagatacagatalaataattgcaaatgagtccttgagatattataaata 4164
DB 4369 AAAGATCAGTGAATGATATTAATTTGACAGGAAGATTAATTAATTAATTAATTAATTAAT 4428
OY 4165 ttatccgaaaataataatcagattgagattcaatlaaaaaataatacaac----- 4215
DB 4429 GTTTTAGCGAAATATAGGATGATTTAGATCAATTAATAAAAGTTATCAAAAGAGAAAG 4488
OY 4216 -----gacaaagaa 4224
DB 4489 GAGAGTTCCCATCATCACACCAACACACCTCCGTCACCGACGAAAAACAGCGAACAA 4548
OY 4225 ggtgaaaatgagaatlacctccctttaaacaatlatgagacattalataaacaagt 4284
DB 4549 AAGAGAGAAAGTAATTCCTTCATTTTAAACAACATTTAGAGACCTTATACATTAACCTTA 4608
OY 4285 aatgataaatgatttatttgaattcattagaagaagaagtcataatatacatat 4344
DB 4609 GTTAATTAATTAATGACATTTACTTTAACTTAAGGCAAAAGATTAACATTAATTAATGTT 4668
OY 4345 gagaatcaaacgtagaagttaaataaagaacttaacttaataaacaatcaagac 4404
DB 4669 GAAAAAGATGAGCAGCATTTAAATTAACCTTAAGATTTTAAAGCAATTTGATGAC 4728
OY 4405 aatgtgcagatttataaaaaataacaatttcgttggaattgctgattatatacaagat 4464
DB 4729 AAAATAGATCTTTTAAACACTAAGCATGCTGGAAGCAATTAATAAAATTTGATTAATGAT 4788
OY 4465 tataaccataaacttatgacaaagtccttagtacaggtatggttttgaataatctt 4524
DB 4789 GATACGAAAAAAGATATGCTTGCGCAATTAATTAAGTACAGGATGAG---TTCAAAATTTT 4845

```

```

OY 4525 gctaaaccgtttatcatatcttaattactgtaggaactgtgcaaggtatgttaaacattca 4584
DB 4846 CCTAATACATTAATATATCAAAATTAATGAAGCAAAATTCACAGATATGTTAAACATTTCA 4905
OY 4585 caaccacaaatgctaaaaaaaacaatgtccacaacaatctcgtatgttccagacattagat 4644
DB 4906 CAACACCAATGGGTAAATAAAACAAATGTCAGAAATTTGATGATGTTGACATTAATGAT 4965
OY 4645 gaaagaagaatgttaaatgtttatlaaattacaacaagaagtgataatggtgtgaa 4704
DB 4966 GAAAGAGAAAGATGTAAAGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5025
OY 4705 aatccaaatcctactgttagcagaataatggtgtagatgtgagatgtgcgaatgtacc 4764
DB 5026 AATCCAAATCTACTGTGTAGCAAAATTAATGAGGTGATGTGATGATGATGATGATGATGAT 5085
OY 4765 gaagaagatcagtagcaagaagaataacacatgtagatgtlaactlaaactgattct 4824
DB 5086 GAAAGAGATTCAGGTAGCAGCAAGAAAGAAATCAGATGGAATGATCAATCACTGATTCCT 5145
OY 4825 tatcaactttcagatgtatttctgcagttcccttaactctttaggaatcatctcta 4884
DB 5146 TATCCACTTTGATGATGATTTTTCGACGTTCTCTTAACCTCTTAGAATATCATTCCTTA 5205
OY 4885 ttaatactatgtlaatatatacatagttcatcttaa 4920
DB 5206 TTAATACATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5241

```

```

RESULT 13
PFAMMSA
LOCUS PFAMMSA 5754 bp mRNA INV 14-MAR-1994
DEFINITION P.falciparum major merozoite surface antigen (PFMSA) mRNA, complete cds, isolate FC27.
ACCESSION M19143
VERSION M19143.1 GI:160412
KEYWORDS antigen; glycoprotein.
SOURCE P.falciparum (isolate FC27 from Papua New Guinea), cDNA to mRNA, clone Ag75, gi.1, gi26, pep3.3.
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE Peterson,M.G., Coppell,R.L., McInlyre,P., Langford,C.J., Woodrow,G., Brown,G.V., Anders,R.F. and Kemp,D.J. Variation in the precursor of Plasmodium falciparum Mol. Biochem. Parasitol. 27, 291-302 (1988)
TITLE JOURNAL MEDLINE FEATURES
SOURCE 88142999
Location/Qualifiers
1..5754
/organism="Plasmodium falciparum"
/isolate="FC27 (Papua, New Guinea)"
/db_xref="taxon:5833"
<1..5754
/gene="PFMSA"
1..5754
/gene="PFMSA"
121..177
/gene="PFMSA"
121..5226
/gene="PFMSA"
/note="precursor"
/codon_start=1
/product="major merozoite surface antigen"
/protein_id="AA029653.1"
/db_xref="gi:160413"
LIFEPROF:SPLEFIINTOCVTHESYOELVKLELBDPAVLNGS
LIFEPROF:NGTSGTAVTSTPGSSGVSVTGSVAASVAASGSGSVASGSGS
RTNPSDMSDSNRTKYADLKHRYONTLETTKELYPPLPULTRNMLTSLNVDGFKY
LIDGEEINELLYKINFYDILRAKLINACANSYCOIPFNKIRANEDLDVLYKTYGQ
RKPDPNIDKNYDMEDYTKRNKTYIANINELIDGSKTIDONKNADNEGRKKLYGQ
YNLFYKQLOEAHNLISVLEKRIDTLKRNINIKLLIEDIDKITDAENPTGSPNP
LPENKKRVEBHEEKIKRIANTIKFNIDSLFTDPLELEYLYREKNKKKVDYPRKSDPT

```


Db	2860	CAAAACCAACATCAATGCATCCTCTACCAATACC	CAAAATGCTGTAAGCTGATCATCT	2919
Oy	2718	-----agtaagtcgaatgtagtaacatacattca	caaatltagaatagatg	2766
Db	2920	GGTCTCTGCTGATGTGGAAGAAAGTCAATGATC	CTTTACAGCTATTGCTATTAGTAACAT	2979
Oy	2767	ctaaatatttggaaaacatctgagctctgaaaaa	acaaatatalataccaagaatta	2826
Db	2980	TTGGAAGAGTATGTTAGTCTCTTAATCTTGA	AAATTAACGTAAAGTACCTAATCCATTA	3039
Oy	2827	---ataggtcaaaaagtagtgaacaccttatalg	aaagatcttaaagaatgtyata	2883
Db	3040	ACCAATTTCTACAAACAGAGATGAGAAAAAT	TTTAGCAAAATTTTAAAAAATTAATGATACC	3099
Oy	2884	ttttataataatcttcttaaaaaatttgtaaaa	acgcgatgatataatcatctg	2943
Db	3100	TATTTTATGTATGATATCAACAAATTCGTAA	ATTCATTTCAAAAGTAAATTTCACAGTTTG	3159
Oy	2944	aatgatatacaaaaaggaagaatctagaagaat	atattatnaatctlaaaaaaacctta	3003
Db	3160	A---CCGAACACAAAAAATGCAATTAAATGA	TGATGAATTTAAAAATGATCTTTA	3216
Oy	3004	cagttatcaatttgatttataataataataaat	aaatctgaagaagattctgtataa	3063
Db	3217	CAGTTATCATTTGATTTTATTAATTAATTA	ATTAATTAATTTGATGATTAATTTAATAG	3276
Oy	3064	aagaacaagttgtgtaataataaatgcaaat	taaaaaaaccttactttattaaagaaca	3123
Db	3277	AAAAAGAACTTGCCCAAGACAAATAATGCA	ATTAAAAAACCCTTACTTTATTAACAAACA	3386
Oy	3124	ctagaatcaaaatggaattcaacttaaaccca	aaagcatgtaatacaaaacctctgt	3183
Db	3337	TTAGATATCAAAATTTGATTCACCTTAAT	TAATTAATGATGATTTAAAAAAGCTTTCTG	3396
Oy	3184	ctctttaacaaaaaagaagctgtgaatagca	gaacactgnaaacacattagaaca	3243
Db	3397	TTCTTTAAACAAAAAAGAAAGCTGAAAT	TAGCGAAGCTGAAACACATTAGAAAAACA	3456
Oy	3244	aaaattattatgaaacctttaaagaactgt	taaatattataatgggtaacatctcca	3303
Db	3457	AAAAATATTATTGAACCTTTTAAAGACCT	TTGTAATTAATTAATGTCGAATCATCTCCA	3516
Oy	3304	ctaaaaccttlaagtgagaatcaattccaac	agaagataactatagccagtttagaaac	3363
Db	3517	TTAAAACTTTAATGTGAATGATATTCAT	TTTCAACAGAAAGATTAATATGCCAATTT	3576
Oy	3364	ctaaaglatthaagtaaatagagaanaaat	taagaatacttaaatctagaaga	3423
Db	3577	TTTGTAGACATTAAGTAAATAGATGAAAA	CTCAATGATTAATTTACATTTAGCAAA	3636
Oy	3424	aaattatcatcatatcaatgtagatcacat	cttattatgctgcgaattaaagaatata	3483
Db	3637	AAATTAATCTTTCTTAATCAAGTGAATTA	CACTTTTAATTAAGTGAATTTAAAAAG	3666
Oy	3484	aaaaataaaaatctatacaggttaattcc	aaagtgaanaataatagogaatgtaaca	3543
Db	3667	AAAAATTAATAATTTACAGGTAATTTCT	CCAAGTGAATAATTAAGAAAGTTTAC	3756
Oy	3544	ctagaatcttacaanaaaattctccaga	agaagacagat-----	3582
Db	3757	TTAAATATCTTACGAAATTTTCTCC	GAAAGCAAAAGTTTACAACAGTTGTA	3816
Oy	3583	-----gttcgaacagtgctgtaagtga	agtgtgataccgaacatlagaa	3624
Db	3817	CAACAGATGTAACTCCATCTCCATTTAT	CTGTAAAGGTAAAGTGAAGTTCCAGAT	3876
Oy	3625	caaaagtcacccaagaacacagcatcaact	catctgtagagagagagttctaaacaa	3684
Db	3877	AAAAAGAAACACAAATTAACCACTTA	AGGCTCTTTATTACGAATTTACACAA	3926
Oy	3685	acatcacaanaatgctgatagtgaatgata	gtgataatcatagtaactatatttgagaa	3744

Db	3937	CATTACCAAAATTTTGACGAGGAAGATGATCTCTTGTAGTTGTATATACCAATTTTGTGCACA	39986
Oy	3745	tcgagaagaagatltatgtatgtlttaggacaagttagtacaaggaagcaagttaactctcc	38044
Db	3997	TCCGAAAGTAATATGACGAATATTTAGATCAAGGTAGTAAGTCGGAAGCAATA---TCTGTC	40534
Oy	3805	gtaatlgtacaactacttcttaaatltgaanaatgaatagaatgaggtttatattttaaacct	38644
Db	4054	ACAAATGATATATCTCTGACGATTTTGAAGAAATGAAATATGATGTATATATTTAAACCT	41133
Oy	3865	ttgacagatgtcttatagaagttttaaanaaacatltagaanaatgaatlttagcatttat	39244
Db	4114	TTAGCTGGAGTATATACAGCTTTAAAAACCAATTGAAAAAACCTTTATTKATTTTAT	41733
Oy	3925	gttaatgtlaaggaatlttlnaaatltcaagatlttaataaagltgaanaatltcaanaatgt	39844
Db	4174	TTAAATTTGAACGATATCTTAAATTCACGCTTAAGAAACGAAATATTTCTTACATGTA	42333
Oy	3965	ttgaatcagaatttaattccatatagaatttaacaatgaataltatgtgtcagaagat	40444
Db	4234	TTAGAACTGTATTTAAATGCAATTTAAACATATATCTCCAAATGAATATATTTGAAAGAT	42933
Oy	4045	ccataaattctctataagaanaaagagataaattcttgaagatlttaaatat	41044
Db	4294	TCATTTAAATATTATGAAATTCAGACAAAAAACACTTTTAAAAAGTTACCAATATATA	43533
Oy	4105	aaggattcoatagaatagogaatataaattltgcgaatgaatgtctctgtgatataataa	41644
Db	4354	AAAGATCAGTAGAAGAAATGATATTTAAATTTGCACAGGAAGGTATAGTTATATGCAAAAG	44133
Oy	4155	ttatccgaaaataataaaatcaagatttagattcaatttaaaaatataaac-----	42155
Db	4414	GTTTATGCGAAATATTAAGATGATTAGATCATTTAAAAAGTATACAAAGAACGAAG	44733
Oy	4216	-----gacaanaa	42244
Db	4414	GAGAGTTCCATCATATACACACACACACACCTCGCTCACGACAAAAACGACGACAA	45333
Oy	4225	ggtgaanaatggaanaataccttcccttlttaaaacatatatgagacctatataaacglt	42844
Db	4534	AAGAAGGAAGAACTAGTCTCTTCCATTTTAAACAAATGTAGACACCTTTATCAATTACTTA	45933
Oy	4285	aatgataaattgtatttattgttaattcaatttagaagcaaaagtcctaattatacatat	43444
Db	4594	GTTTATTAATTTGACGATTCCTTAATTAATCTTAAAGCAAAACATTAAACGATTTGTAATGTT	46533
Oy	4345	gagaaatcoaaacgtgagaagtttaaaagaacttaattactttaaaacaatcoaaac	44044
Db	4654	GAAAAAGATGAGCACACTGTTTAAATTAATTAACCTTAGTGATTTTAAAGCAATTTGATGAC	47133
Oy	4405	aaattgycagattttaaanaaaataaacaatttcgttgaatgtgcgattatcaacagat	44644
Db	4714	AAATATAGATCTTTTAAAAACCTAAGCGACTTGAGCAATTTAAAAATGTATAATGAT	47733
Oy	4465	tataacataataacttatgtgacaaagttcccttagtaagatagaagtttltgnaaactt	45244
Db	4774	GATTCGAAAAAAGATATGCTTTGGCAAAATTAATCTTAGTACAGGATTAAG---TTCACAAATTTT	48330
Oy	4525	gctaaacgcttatttcttaattacttgaatgttgaactgtcaagtgatgtttaaactttca	45844
Db	4831	CTTATATCAATTAATATCAATTAATTAATTTGAAGGAAATTCACAGATATGTATTAACATTTTCA	48930
Oy	4585	caacacaaatgtcgttaanaaaacaatgtccacaanaatlttcgaatgttltcaagcaattgat	46444
Db	4891	CACACCAATGCGTAAAAAACAATGTCAGAAAAATTCGTGATGTTTCAGACATTTAGAT	49530
Oy	4645	gaagagaagaatgtataaagtlttataaattacaacaagaagaagtgtataaagtlttgaa	47044
Db	4951	GAAAGACAGATATTAATGTTTATTAAATTAACAACAGAGGTGATTAATATGTTCTGAA	5010
Oy	4705	aatccaataacttactgttaacgagaanaatagtgtgatatgtatgcgaatgtcaaatgtcac	47644
Db	5011	AATCCAATCTTACTTGTATAACGAAATATATGTGGATGTGATGACGAGTCCACATGTATGC	50700

OY	4765	gaagaagattcagcgtagcaacggaagaaataatcacatctgtgaatgtactacaacctaatctt	4824
Db	5071	GAAAGAAGATTCAAGGTACGACGAGAAAAGAAAAACATCATGTGAATGTAACAACTCATTTC	5130
OY	4825	taccacatttcgaatgtattctgcagctcccttaactccttaggaatcatctcta	4884
Db	5131	TATCCACTTTTGCATGGATTTTTTCGCAGTTCCTCTAATCTCTTAGGAATATCATITCTTA	5190
OY	4885	ttaatactactgttaataatatatcacgtttcata	4920
Db	5191	TTAATACATCATGTATAATATTATACAGTTTCATTTTAA	5226
RESULT	14		
PMEZSAIB			
LOCUS	PMEZSAIB	5312 bp	DNA
DEFINITION	P.falciparum gpi190 (MSAI, MSP1, PMMSA)	INV	10-MAR-2001
ACCESSION	235327	U09209	merozoite surface antigens.
VERSION	235327.1	GI:929795	
KEYWORDS	gpi190 gene; merozoite surface antigen; MSAI gene; MSP1 gene; PMMSA gene.		
SOURCE	Malaria parasite P. falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 5312)		
TITLE	Pan,W., Tolle,R. and Bujard,H. A direct and rapid sequencing strategy for the Plasmodium falciparum antigen gene gp190/MSAI		
JOURNAL	Mol. Biochem. Parasitol. 73 (1-2), 241-244 (1995)		
MEDLINE	96123395		
REFERENCE	2 (bases 1 to 5312)		
AUTHORS	Tolle,R., Bujard,H. and Cooper,J.A.		
TITLE	Plasmodium falciparum: variations within the C-terminal region of merozoite surface antigen-1		
JOURNAL	Exp. Parasitol. 81 (1), 47-54 (1995)		
MEDLINE	95354793		
REFERENCE	3 (bases 1 to 5312)		
AUTHORS	Tolle,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (18-JUL-1994) Ralf Tolle, Abt. Prof. H. Bujard, Zentrum fur Molekulare Biologie, (ZMBH), Univ. Heidelberg, Im Neuenheimer Feld 282, Heidelberg, 69120, Germany		
COMMENT	On Aug 5, 1995 this sequence version replaced gi:535249.		
FEATURES	Location/Qualifiers		
Source	1..5312		
	/organism="Plasmodium falciparum"		
	/isolate="NF54"		
	/db_xref="taxon:5833"		
	/chromosome="9"		
	/clone="3D7A"		
	/dev_stage="blood stage"		
	109..165		
	/gene="gpi190, MSAI, MSP1, PMMSA"		
	109..5271		
	/gene="gpi190, MSAI, MSP1, PMMSA"		
	109..5271		
	/gene="gpi190, MSAI, MSP1, PMMSA"		
	/codon_start=1		
	/product="precursor of the major merozoite surface antigens"		
	/protein_id="CA84556.1"		
	/db_xref="GI:929796"		
	/db_xref="SPTREML:Q25922"		
	/translation="MKTIIFPCSLFFIIINTQVTHESYOELVKLEALDAVLTVGS LFQEKVNLNEEITTKSGASOASGASOASGASOASGASOASGASOASGTSGP SGPGTSPSSRSNTLPKPSNTPSGASPADAASDAKSYADLKHVRVLYPTIKLKYP EIPIPTAMLTIDNIIGDKFYLLIDGYEINIELLYKLAFYDLIPAKINDVCANCOI PENLKTANMELDYLAKLVFGYGRIVNDIKDNVGMEDIYKNKTTIANINBLEGSKR PTINDKANADEEGKRRLYQAQVDISLYNKOLEEHNLISVLEKRIIDLKNNENIRKL GLKNEINPPANGNPNTPLLNKKITEEEHEEKIKETIKFNIDSLETNPDELEY YLEKNNKVVTPEKSOPSTSVQIPKYVPNGIYPLPLNDIHNSLAADNDKNSGDL		
sig_peptide			
gene			
CDS			

MPHTKEKINEKIITDNKERKIFINNKKIKIDLEKNINHTKEONKKLEDEYKSKKD
YEELLEKEFEYEMKNNPDKVDYDKJFSAYTVNVKORNNKPNSSNNVYVOKLAK
ALSYLEDESLRKISSEKDNHYTLKTGLADIKLETEIKSEKKLENNKNGKGLTHS
ANGSLESDIVKLOVOKYLIKIDRLKIELFKLELNAOLKDSHTNPNITKPNKPEPY
YLIIVKKEVDKKEFIPKYKDKLKKKEQALSSITOPPLVANSSTTEDGGSTHTLSOG
ETEVEETEETEVEYGHHTVTTPPLPPOPSPEKKEVENSIEKSNNSOALTKRTV
YIKRIDEFLTKSYIKHYILVNSSMDOKLLEVYNLTPREENELKSCDPLSDLFNPN
NIPAMYSLYDSMNNDLOHLFELLYOKEMLYLHKLEENHIEKLEEQOIGTSTSS
SPGNVTVAOSATSHNSNOOSNASSTONGVNSDPAYVESHPDLJVLJSTSDND
LKGISVLNKGKTKVNPDLTSTTEMEKEFNILKNDTNPNDNDIKOVKNSVYIT
CETEPKNAINDEIKKIDLOLSFDLYKRYKLIKDRLEPNKKEIGDOMKRIKLITL
KQDLSKLSLNNPNHVLONEVFPNKKKEALIEAETENLENTKILKRYKIDLVYYN
GESSPLKTLSEVSIQTEDNYANLEKRVLSKIDGLNNDLHLGKKLSLSGLHLI
TELKEVIRKNNKNTGNSPENNNKVNKAJSYENLPKAVTVTVPPQDPVPSPLSV
RVSGSSGSKETEOJPTSGSLLELOOVYONVDEEDSLVYVLPJEGSEDNDELD
OYVGEAISVTMDNLTSGFENEDYVYIKPLAGVRSJKOIEKIEFTFNNLNDILN
SLKKRKYFLVYEDSLDMOFKRIISSNEYIIESFPLNSSEOKNTLKSRYIKESVEN
DIKFAOEGISYIEKVLAKIDLESITKRYKEKKEKFPSSPTTPPSKPTDEOKES
KFLPPLTNETLYNNLVNKIDYLIINLAKINDCNVEKDEAHVKITKLSDLAIDKI
DLKPNPYDEFLIKRLINDTKKMDLGLSTGLVONPPTIISKLEGFODMLNISO
HOCVKOCPENSGCCFRHLDERECKLNLKOEGRKVCVENPPTCNENNGGCDADATC
TEEDSGSSRRKKTCECTKPDYVPLFDGIFCSCSSNPLGISLILMLILXSF"

BASE COUNT 2328 a 703 c 705 g 1576 t
ORIGIN

Query Match 55.2% Score 2713.8 DB 96 Length 5312;
Best Local Similarity 73.3% Pred No. 11e-243;
Matches 3791; Conservative 0; Mismatches 1117; Indels 267; Gaps 14;

OY 1 atgaagacataatcttcttaatgtaattcttcttcttataataacaaatgta 60
DB 109 ATGAAGATCATATCTTTTATGTCATTTCTTTTATATAATACACAATGTGA 168
OY 61 acacatgaagaattataagaacttgcacaaactagaagcttagaagatgta 120
DB 169 ACACATGAAGAATTATCAAAAGCTTGAAGCTTTAGAACGTGACGATTTG 228
OY 121 acaggtatagttatttcaaaaggaataatgtaataagaagaacagtgta 180
DB 229 ACAGGTATAGTTTATTTTAAAAAGAAAATGTAATTAATGAAGAAGAAATTA 288
OY 181 gctgtacacactagtaacactgttcaaaaggttcagttcaggtggtcagtg 240
DB 289 AAGGTGCAAGTGTCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTG 348
OY 241 tcaagttcaggtggtcaggtcagttcaggtggtcaggtcaggtggtcaggt 300
DB 349 GCAAGTGTCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGT 408
OY 301 gcttcaggtggttcaggttaattcaagac----- 328
DB 409 GGTCCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCA 468
OY 329 -----glacaatcctcaagaataatcaagttatcagatgctcaatctac 375
DB 469 AATTAATCTATCTGTGTCAGACCCCTCAGCTGATGCAAGCGATTCAGATCTTAAT 528
OY 376 gctgatttaaacacacagagtaacaaactgttaactacatacaagaactca 435
DB 529 GCTGATTATAAACACAGATGCAAGTGTGCTTGTCTATTAATAAACACCAATATCC 588
OY 436 caactcttgatttaactacataatgtaactttgtgataataatlaatgattca 495
DB 589 GAACCTTTGATTTAACCATATATGTTAACTTGTGTGATTAATATGATGTTTCA 648
OY 496 tatttaattgagatagataagaataatcaatgataatataataaacttatt 555
DB 649 TATTATAATTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 708
OY 556 gatttataatgaagcaaatataatgataatgataatgataatgataatgata 615
DB 709 GATTATAATGAAGCAAAATTAATGATGATGATGATGATGATGATGATGATGATG 768

OY 616 aatcttaaaatcgtgcaaaatgaatagacgtacttaaaaaacttggttcgata 675
DB 769 AATCTTAATAATTTGTCGAATGAAATTAAGACGTCTTTAAAAAACCCTTGTTGATATAG 828
OY 676 aaacattagacaataataagaataatgtagaaaaatggaagaattacataa 735
DB 829 AAACATTACACAAATTTAAAGATTAATGTGAGAAAATGGAAGATTAACATTA 888
OY 736 aaaaaaacatagaaaaataataatgaatgaatggaagaatgagaagaacat 795
DB 889 AAAACAACATACCAATATAAATGAATTAATGAAGGAAGTAAAGAAACAAATGATCA 948
OY 796 aataagaatgcaactaagaagaagaagaagaagaagaagaagaagaagaaga 855
DB 949 AATAAGAAATGACATTAATGAAGAAAGGAAAAAAATTTATACCAACTCAATATGATCT 1008
OY 856 tctattacaataaacaatagagaagacacataatlaataagcgtttaga 915
DB 1009 TCTATTACAAATTAACAAATTAAGAAAGCACATTAATTAAGCGTTTGA 1068
OY 916 attgacactttaaataaaatgaacatgaagaattactgataagaataat 975
DB 1069 ATTGACACTTTAAAAAATAAGAAACATTAAGAAATTAATGATTAAGATTAATG 1128
OY 976 aaaaatccccaccgccaattctggaataacacgaataactccttgataaga 1035
DB 1129 AAAAATCCCCCAGCGCCCAATTTGGAATTAACCAAAATTAATCTCTTGTGAAGAACAA 1188
OY 1036 aaaaatcaggaacacgaagaagaagaataaagaatgcaaaactlaaattla 1095
DB 1189 AAAATGAGAGAACACAGAGAAAATTAAGAAATTTGCCAAATCTTAATTAACATT 1248
OY 1096 gatagttattatgatacactgtaactgaatgaatactattlaagagaaga 1155
DB 1249 GATAGTTTATTAATCTGATCCACTTGAATTAAGATTAATTAATTAAGAAAAA 1308
OY 1156 attgatatagtgcaaggttga-----acaagaatcaactgaaccaatga 1206
DB 1309 GTTGATGTAACTTAATTAATCAAGATCCACGAAATCTGTTCAAAATTAACAAAGTTCC 1368
OY 1207 tatccaatggaatlaactatccttgatataagaatataaactgctta----- 1260
DB 1369 TATCCAATGATGATTTATATATCTTTACCACTGATGATTAATTAATTAATTAAT 1428
OY 1261 ---aatgaacttaattccttggtgatttaattcaatcattgatt 1317
DB 1429 GATTAATGATTAATAATTAATGATGATTAATTAATTAATTAATTAATTAATTAAT 1488
OY 1318 agtaaaacatatatactgataatgaagaagaagaagaagaagaagaagaaga 1377
DB 1489 GAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1548
OY 1378 attaaaatgaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1437
DB 1549 ATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1608
OY 1438 tcttaaatgataataaagaagaataatgaagaatlaacttaatgaagaat 1497
DB 1609 GATTATGAAGAAAGCAAAAGATTAATGAAGATTAATGAAGATTAATGAAGATTAAT 1668
OY 1498 ttaaatataatataatgaatgaatgaagaagaagaagaagaagaagaagaaga 1557
DB 1669 TTTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1728
OY 1558 aaagttagaagaacttaacacacataacttgcatactgaagaagaagaagaaga 1617
DB 1729 AATGTTGAAGAAACAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1788
OY 1618 ctggaagaatgaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1677
DB 1789 GTTCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1848

1678 gtgaaaaagaattlaaataatlaaataatlaaagaacaataagaataatgaatctgaa 1737
1849 tctgaaaacagattttatcatctttatcttgcaaacctggcccgaaagctgatatgaaa 1908
1738 accttgctgaaaaatataataaagaatgaagaacagctcttctgaaaaaataatctaaa 1797
1909 aaattttaaacaagaamaataaagaatgagaaacaaaattctagaaaaaaa---ttttaa 1965
1798 gaagaaataaacaacagaatgaataaatttgaagatctacagatctgtaaaagtcaagct 1857
1966 ggcattacacatttcagcaaatgcttcttgaaagatctctatatttgtaaaatttgcagta 2025
1858 caaaagcttatttataatgaacaaatgacgaatlaaaaaagaatcgaatctgattttaa 1917
2026 caaaaagttttatt 2085
1918 aatgtgaattaaacataatatacatgttcccaatcttcaacaacagaataaagaacaa 1977
2086 aatgcacacactaaagaaatgactatgacaaatatttttttttttttttttttttttt 2145
1978 gaacctattatttatttctgttgtaaaaaagaataatgataataatgaatgcttcagct 2037
2146 gaacctattatt 2205
2038 aagtgagaatcattgataaagaataaataaataataataataataataataataataata 2091
2206 aaagtttaaaacagatt 2265
2092 tctgataatctcggaacacatcaacgaagaataaatacagaagaacacacataaact 2151
2266 gtttgacgaacagcaac 2325
2152 ggaacacacagcagatctgcttctgagaagagatctgaatgaagaacacacacacacac 2204
2326 ggaac 2385
2205 --agaac 2262
2386 acgcttaac 2445
2263 gtcccaac 2322
2446 gaaattttcaattac 2505
2323 ctggaac 2382
2506 cttaac 2365
2383 tcaac 2442
2566 tcaac 2625
2443 agcaac 2502
2626 aatgaatttttaaaac 2685
2503 gtaattattctatgttctgaagacacacacacacacacacacacacacacacacacacac 2562
2686 gcattatgatttatttatttatttatttatttatttatttatttatttatttatttattt 2745
2563 tatgaac 2622
2746 tatcaac 2805
2623 ttaac 2676
2806 ttt 2865
2677 tcaatgaac 2717
2866 accgtaaatatttctgcttcaatccgacacacacacacacacacacacacacacacacac 2925
2718 -----agtaagtga 2727

2926 tctctac 2985
2728 aatgaatgatacatcac 2787
2986 agctatgatttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 3045
2788 ttagctcttggaac 2844
3046 ttaatttcttggaac 3105
2845 gaaac 2904
3106 gaaac 3165
2905 aatttgaac 2964
3166 caatttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 3222
2965 aattgaagaagac 3024
3223 gcatttaatttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 3282
3025 aataac 3084
3283 aatttaatttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 3342
3085 aaaaac 3144
3343 aaatttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 3402
3145 cttaac 3304
3403 cttaac 3462
3205 gctgaac 3364
3463 gctgaac 3522
3265 aaaaac 3324
3523 aaagac 3382
3325 tcaac 3384
3583 tcaac 3642
3385 gaaac 3444
3643 gatgaac 3702
3445 ggaac 3504
3703 ggaac 3762
3505 aatttgaac 3564
3763 aatttgaac 3822
3565 ctccgaac 3585
3823 ctccgaac 3882
3586 gaaac 3645
3883 ccatttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 3942
3646 gcaac 3705
3943 acttgaac 4002
3706 gaagtgaac 3765

Dh 2038 CATGTACCAATATTTATTAACCAAAATTAACAGAACCATATTTATTTATTTGATTATA 2097
Qy 2002 aaaaaagaatttgataaatttaaaagtgctcagccaaagtgtaagatcattgataaattgaa 2061
Db 2098 AAAAAAGAGTATTAATTAATTAAGATTTATTAACAAAAGTAAACATATGTTTAAAGAAA 2157
Qy 2062 gaaaaaanaaacaataaagacagaagtgatcagtaatttcggaacacaaacagagga 2121
Db 2158 GAACAAGCTGTCTTATCAAGTATTACACAACTTTAGTTGACAGACGCAAC--AACT 2214
Qy 2122 gaataacagagcaagcaactcaaaaactggaacaaagcagatctgctttagaaga 2181
Db 2215 GAGATGGGGCTCCTCCACACACATTTATCCCAATTCAGAGAAACAGAGTAACAGAA 2274
Qy 2182 gattagtgacagcacaagcacaagacaaaacaaagcacaaccagtlaccagtaaca 2241
Db 2275 GAACAGGAAGAAACAGAAACAGTAGACACACAAACGGTAACTAATTAACATTACCA 2334
Qy 2242 gtaacagaagcaaaagcacaagtcaccaacacacagcaaccagtaataaatactgaa 2301
Db 2335 CCAAAAGAGTAAAGTGTGAAATTTCAATTAAGACATTAAGGTATATGCAATTCACAA 2394
Qy 2302 aatgttccaatttagatccttgaaaaatataatgaattttaaactcatalata 2361
Db 2395 GCGTTGACAAAAACAGTTTATCTAAGAAATTAAGATGAATTTTAACTAATCATATATA 2454
Qy 2362 tctgcaaatatatttggcttccaactcaactatgaagaaagatataaacaatat 2421
Db 2455 TCGCATTAATTAATTTTATCTATCAACCTAGATAGACCAAAATTAATTAAGGTATAT 2514
Qy 2422 aaaaatacaagaaggaagcaaatgaatgaatcagtcagtcacatagactatgctt 2481
Db 2515 AATTTTACGCGAAGAGAAATGAATTAATTAATCATGATGATCATTTATTTATTTATTT 2574
Qy 2482 aatatacaaatataacatccgtaactgtaactgtaactgtaactgtaactgtaact 2541
Db 2575 AATATTCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2634
Qy 2542 tcaacactatattgaagaatttatgaagaagaatgcttgtaattatataaacttaag 2601
Db 2635 CAACATCTCTTTTTCATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2694
Qy 2602 gataatgacaaaatttaaatattatagagaagcgaaaaaa-----gtatcacact 2655
Db 2695 GAGGAAATACATCAATAAATTTATGAGAGCAAAACAAATTAATGGAACATCATCT 2754
Qy 2656 gtaaaaaactcttcaagtcacatgaacacatcatcattcaaccctcaagaataacc 2715
Db 2755 ACATCCAGTCTGGAATTAACACCGTAATTAATTAATTAATTAATTAATTAATTAAT 2814
Qy 2716 ga----- 2717
Db 2815 CAAACCAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2874
Qy 2718 -----agtaagtgcaaatgtaatacatcactcactcactcactcactcactcact 2766
Db 2875 GGTGCTGCTGTAAGTAAGAAAGTATGATCCCTTAACAGTATTTGCTATTAGTACGAT 2934
Qy 2767 ttaaaattcttgaaacataatgcttggaaaaaacaaaatataatacagaata 2826
Db 2935 TTGAAGGATTTGTTAGTCTTTAAATCTTGGAATAAATACTAATTAATTAATTAATTAAT 2994
Qy 2827 ---ataagtcataaagaatgtaaaactttatgaagaatataaagaatgtaata 2883
Db 2995 ACCATTCTTACACAGATGGAATAATTTTATGAGAAATTTTAAATAATTAATTAATTAAT 3054
Qy 2884 ttcttaaatgaatcttcttaaaatttgtaaaatcctcaagcctgataatcaatcact 2943
Db 3055 TATTTTAATGATGATTAACAAATTCGTAATAATTAATTAATTAATTAATTAATTAATTAAT 3114
Qy 2944 aatgatgaatcaaaaagaagaattagaagaatataataaataaataaactta 3003
Db 3115 A---CCGAACACAAAAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3171

Qy 3004 cagttactttagattatataataataaataaataaataaataaataaataaataaataaataa 3063
Db 3172 CAGTTATCATTTGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3231
Qy 3064 aagaacacagcttgtaataataaataaataaataaataaataaataaataaataaataaataa 3123
Db 3232 AAAAAAGACTTGCCAGACGAAATGCAAAATTAATAAATACTTATTAATAAAGAACAA 3291
Qy 3124 ttagaactaaatttgaaacttaacttaataaataaataaataaataaataaataaataaataa 3183
Db 3292 TTGAATCAAAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 3351
Qy 3184 ttcttcaaaaaaataaagaatgtaagaagcaagcaagcaagcaagcaagcaagcaagcaagca 3243
Db 3352 TTCTTTTACAAAAAAGAGCTGAATTAAGCAAGAACTGAAACATTAAGAAACCA 3411
Qy 3244 aaataataattgaaacattataaagacttgtaaatatataatgtaactcaca 3303
Db 3412 AAAATTTTATTTGAACATTAATAAGGACTTGTAAATTAATTAATTAATTAATTAATTAAT 3471
Qy 3304 ttaaaaaacttgaagtgaaatcaactcaaacagaagaataatgacagttagaacac 3363
Db 3472 TTAAAACTTTAAGTGAAGTATCAATTAACAGAGATTAATTAATTAATTAATTAATTAAT 3531
Qy 3364 tttgaagttatgaataattagaagaagaataaagaatgaatgaatgaatgaatgaatgaat 3423
Db 3532 TTATGAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3591
Qy 3424 aaatataacttaactcaagtgatgatacatcattatgcttgtaacttaagaagaataa 3483
Db 3592 AATTTTCTTTCTTATCAAGTGTATCATCTTATTAATTAATTAATTAATTAATTAATTAATTA 3651
Qy 3484 aaaaataaataatataacagtgtaacttccaagtgaaataaataaataaataaataaataa 3543
Db 3652 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3711
Qy 3544 ttagaactcttaaaaaatttctcccaagaagaagaagaat----- 3582
Db 3712 TTAAATCTTACGAAATTTTCTCCAGAAACGAAAGTTTACAACTTTACTCCACT 3771
Qy 3583 -----gtgcaacagcttgtaagtgaaagtgagtcgacacatlagaa 3624
Db 3772 CAACGATGTAATCTCATCTCATCTCATCTCATCTCATCTCATCTCATCTCATCTCATCTCAT 3831
Qy 3625 caaagtcaaccaaagaacacagcaactcaactcagtgagcagagctcaacataaaca 3684
Db 3832 AAAGAGAAACCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3891
Qy 3685 acatcaaaaatgctgataatgagtagatgagtagatgagtagatgagtagatgagtagatgag 3744
Db 3892 CAATTTCAAAATTTATGACGAGAGATGATCTCTTATTAACAGAAATTAACAAAGTAGTA 3951
Qy 3745 tccgaagaagattatgatatattaggaacagtagtaacaggaagaagcaglaactccttc 3804
Db 3952 TCCGAAGATTAATGACGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4008
Qy 3805 gtaaatgataacacttcttaaaattgaaatgaatgaatgaatgaatgaatgaatgaatgaat 3864
Db 4009 ACATGATGAATTAATCTCTCAGAGATTTGAAATGAATGAATGAATGAATGAATGAATGAAT 4068
Qy 3865 ttagagcgtgattatgaagaatgtaaaaaaataaataaataaataaataaataaataaataa 3924
Db 4069 TTAGCTGGAGTATTAATGAAGCTTAATAAATAAATAATTAATAAATAAATAATTTTACATTAAT 4128
Qy 3925 gtaaatgtaaggaatatttaaatcaagatataaataaataaataaataaataaataaataa 3984
Db 4129 TTAAATTTGACAGATTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4188
Qy 3985 ttagaactcagatttaactccatataaagaattcaacatcaaglaattatgctgcaagat 4044
Db 4189 TTGAATCTGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4248

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2001, 15:51:20 ; Search time 53.94 Seconds
(without alignments)
4020.169 Million cell updates/sec

Title: US-09-269-874-3
Perfect score: 8424
Sequence: 1 MKIIFLCSEFLFIINTOCV.....SNFLGISFLIMLILYSFI 1639

Scoring table: BLOSUM62 , Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP:REMBL_16:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.podent:*
13: sp.spine:*
14: sp.virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	5283.5	62.7	1704	5	09TZT4	09TZT4 plasmodium
2	5162.5	61.3	1694	5	09TZT5	09TZT5 plasmodium
3	5162.5	61.3	1694	5	09NHX1	09NHX1 plasmodium
4	5107.5	60.6	1720	5	025922	025922 plasmodium
5	3093.5	36.7	1751	5	026194	026194 plasmodium
6	3003.5	35.7	1087	5	025961	025961 plasmodium
7	2951	35.0	1726	5	002569	002569 plasmodium
8	2801	33.3	539	5	09TYG1	09TYG1 plasmodium
9	2798	33.2	539	5	025972	025972 plasmodium
10	2796	33.2	539	5	025966	025966 plasmodium
11	2795	33.2	539	5	025973	025973 plasmodium
12	2794	33.2	539	5	025976	025976 plasmodium
13	2791	33.1	539	5	025981	025981 plasmodium
14	2789	33.1	539	5	025984	025984 plasmodium
15	2788	33.1	539	5	025971	025971 plasmodium
16	2532	30.1	1787	5	025645	025645 plasmodium
17	2530	30.0	1791	5	09U6D4	09U6D4 plasmodium
18	2509	29.8	1785	5	025685	025685 plasmodium
19	2494.5	29.6	1766	5	025668	025668 plasmodium

20	2146	25.5	651	5	025924	025924 plasmodium
21	2119.5	25.2	652	5	025923	025923 plasmodium
22	2101	24.9	400	5	003999	003999 plasmodium
23	1818	21.6	360	5	09NAT3	09NAT3 plasmodium
24	1778	21.1	569	5	025975	025975 plasmodium
25	1772	21.0	569	5	025969	025969 plasmodium
26	1772	21.0	569	5	025977	025977 plasmodium
27	1768	21.0	569	5	025974	025974 plasmodium
28	1768	21.0	569	5	025979	025979 plasmodium
29	1767	21.0	569	5	025983	025983 plasmodium
30	1765	21.0	569	5	025967	025967 plasmodium
31	1758.5	20.9	570	5	025968	025968 plasmodium
32	1753	20.8	569	5	025982	025982 plasmodium
33	1749	20.8	569	5	025970	025970 plasmodium
34	1748.5	20.8	570	5	025972	025972 plasmodium
35	1748	20.8	569	5	025978	025978 plasmodium
36	1745	20.7	569	5	025980	025980 plasmodium
37	1696	20.1	336	5	025865	025865 plasmodium
38	1670	19.8	356	5	09TZU5	09TZU5 plasmodium
39	1649	19.6	350	5	09TZU2	09TZU2 plasmodium
40	1636.5	19.4	357	5	09TZU7	09TZU7 plasmodium
41	1634.5	19.4	357	5	09TZU8	09TZU8 plasmodium
42	1625.5	19.3	360	5	09TZV2	09TZV2 plasmodium
43	1625.5	19.3	363	5	09TZU9	09TZU9 plasmodium
44	1622	19.3	344	5	09TZU1	09TZU1 plasmodium
45	1616.5	19.2	362	5	09TZV3	09TZV3 plasmodium

ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	1704 AA.
09TZT4	09TZT4			
AC	09TZT4			
DT	01-MAY-2000 (TREMBLrel. 13, created)			
DT	01-MAY-2000 (TREMBLrel. 13, last sequence update)			
DT	01-JUN-2000 (TREMBLrel. 14, last annotation update)			
DE	MEMOZITTE SURFACE PROTEIN 1 PRECURSOR.			
GN	MSPI.			
OS	Plasmodium falciparum.			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI_TaxID=5833;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=HN2;			
RA	Jiang G., Liu R.Z., Daubenberger C.A., Pluschke G.;			
RL	Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF062349; AAC72885.1; -			
DR	InterPro: IPR000561; -			
DR	Pfam: PF00008; EGF; 1.			
KW	Merozoite.			
SQ	SEQUENCE 1704 AA; 193762 MW; 385526DDA56FD1D CRC64;			
Query Match	62.7%;	Score 5283.5;	DB 5;	Length 1704;
Best Local Similarity	62.1%;	Pred. No. 2.1e-187;		
Matches 1075;	Conservative 219;	Mismatches 317;	Indels 121;	Gaps 26;
OY	1 MKIIFLCSEFLFIINTOCVTHESVOELVKLEALDELAVLIGYSLFOEKWLNIGTSGT 60			
Db	1 MKIIFLCSEFLFIINTOCVTHESVOELVKLEALDELAVLIGYSLFOEKWLNIGTSGT 60			
OY	61 AVTSTPGSK--GSVASGSGSVASGGSVASGGSVASGGSNRRTPSDNS 117			
Db	61 AVTSTPGSKSGSVASGSGSVASGGSVASGGSVASGGSNRRTPSDNS 117			
OY	118 SDSAKSYADIKHVRNVLITIKELYPQLPDLTNHMTLTDNTHGFRYLLDGEETNEL 177			
Db	118 SDSAKSYADIKHVRNVLITIKELYPQLPDLTNHMTLTDNTHGFRYLLDGEETNEL 177			
OY	178 LYKLFYFDLLRAKINDVCANDYCOIPENLKRANEDLVLLKLVFGYRKPNDNKDNNGK 237			
Db	178 LYKLFYFDLLRAKINDVCANDYCOIPENLKRANEDLVLLKLVFGYRKPNDNKDNNGK 237			

```

Db 178 LYKLFYDILLRAKINDACANSYCOIPFNELKIRANELDVLAKKIVGRRKPLDNIKDNVKG 237
Qy 238 MEDYTKKKKTENTENINELIESSKTIIDKNKATKEEKKKIYQAOYDSTYNKOLEEPAHN 297
Db 238 MEDYTKKKKTITIANINELIESSKTIIDKNKADNEGGKKLIYQAOYDSTYNKOLEEPAHN 297
Qy 298 LISVLEKRIIDTLKKNENIKELLDKINEIKPNPPANSNGTPTLIDKNKIEEHEKEI 357
Db 298 LISVLEKRIIDTLKKNENIKELLDKINEIKPNPPANSNGTPTLIDKNKIEEHEKEI 357
Qy 358 AKTIKFNIDSFTDPLELEYLREKKNIDISAKVE--TKESTEPN-EYPPGVTYPLSYN 414
Db 358 AKTIKFNIDSFTDPLELEYLREKKNIDISAKVE--TKESTEPN-EYPPGVTYPLSYN 414
Qy 415 DINNAL---NELNSFGDILNPDPYTKPEPSKNYTDN-ERKKFNEIKREKIEKKI--- 467
Db 415 DINNAL---NELNSFGDILNPDPYTKPEPSKNYTDN-ERKKFNEIKREKIEKKI--- 467
Qy 418 DIHNSLADNKNSTGYDLMPN-DYREKINEKTIIDNKERKIFINNIKOIDIEENINHT 476
Db 418 DIHNSLADNKNSTGYDLMPN-DYREKINEKTIIDNKERKIFINNIKOIDIEENINHT 476
Qy 468 -ESDKSYEDRSKSLNDITTEYKELNEIYDSKFNNNIDLTNEKMGKRYSYVEKLT 526
Db 477 KEONKKTLEDYEKS---KDYEBELLEKFYEMKFNNNPDKDVOKIFGARYTANVEKQRY 532
Qy 527 HNTFASYSKHNLEKLTALKYMEDYSLRNIVAEKLYTKNLISKIENEIETLVENIK 586
Db 533 NKKSSSNNSYVYOKLKLALSYLEDYSLRKGISSEKDNHYTLTGLEADIKLEIK 592
Qy 587 KDEOLFEEK---ITKDNKRPDEKILEVSDIYKVOYOKVLLMNKIDELKKTOLIKNVEL 643
Db 593 SSEKILEKKNKGLTHSNANAS---LEVSDIYKVOYOKVLLMNKIDELKKTOLIKNVEL 648
Qy 644 KHNHVPNSYKQENKQEPYLLVLKEIDKLVPNKVESLINEKKNIKTEGOSDNSE 703
Db 649 KDSIHVPNIYKQNKPEPYLLVLKEIDKLKEFIPKVKDMKKQAVLSS----- 699
Qy 704 STEGITQATTKPQOQGSALIEDSYOAOQOQAO-----PPVPPVPEAAQAP 757
Db 700 ITQPLVASETTEDEGSHTHLSOSGETEVEETEETETEGVHTTAVITLPTP----- 756
Qy 758 PPAPV-----NKTENVSKLDYLEKLEFLNLTSYCHKILVSHSMNKKILKQ 806
Db 757 PPKYKVVYENSTIEHKSNSQALTYVTKLDEFLTKSYCHKILVSNSSMDKLEVL 816
Qy 807 YKITEESKSLSCDPLDLFNIONNIPVMSMPDLSNLSQLEMEIYEMKVCNLYKL 866
Db 817 YNLPEEKNEKSCDPLDLFNIONNIPVMSMPDLSNLSQLEMEIYEMKVCNLYKL 866
Qy 867 KONDKIKKLEBAKV-----STSVKTLSSS-----MPLSTLPDQKPEVS 908
Db 877 KEENHIKRIEOKOITGTSSTSPGNTVTVAQATHSNSQOQSNASTWTONGAVAS 936
Qy 909 AND---DTSHS---TNLNSLKLEFNILSGKNKNITQOL-IGQKSENEFEKILKSD 960
Db 937 SGPAVESHPPLVYLSNDLKGIVSLNKGKTKVPNPLITSTTEKEFEENILKND 996
Qy 961 TFYNSFTNPKSRADILNSINDESKRRLKEDJNLKKTLOLSFDLYNKKYKLEFLR 1020
Db 997 TYPNDNDIQFVKSNSKVITGLT-ETQKNAIODEIKKLDIOLSDLYNKKYKLEFLR 1055
Qy 1021 KKKVYKTKMOIKKLTLEKQLESKLSLNNPKHVLQNFVFPNKKKAEALAEIEN 1080
Db 1056 KKKVYKTKMOIKKLTLEKQLESKLSLNNPKHVLQNFVFPNKKKAEALAEIEN 1080
Qy 1081 TKILLKHKGLVKKYNGSSPLKTLSEESIGTEDVYASLENFVKYLSKLEKYLKOLNLEK 1140
Db 1116 TKILLKHKGLVKKYNGSSPLKTLSEESIGTEDVYASLENFVKYLSKLEKYLKOLNLEK 1140
Qy 1141 KKLSTLSSGLHLTALAEVLKKNKYTGNPSSENNTPDNNALLESYKFLPREGDVA 1200
Db 1176 KKLSTLSSGLHLTALAEVLKKNKYTGNPSSENNTPDNNALLESYKFLPREGDVA 1200
Qy 1201 E-----SGSDTLEQSQPKRPASTHGAESNTITTSQNDVEDVDVITVIF 1246
Db 1235 PPODPVTPSPLSVRSVSGSGSTKETQIPTSGSLTLEQVOVQSONYDEEDSLVLE 1294

```

```

Qy 1247 GSEEDYDILGOVYTGAVTPESVIDNLSKTIENEYEVLYKPLAGVRSKLOENNYM 1306
Db 1295 GSEEDYDILGOVYTGAVTPESVIDNLSKTIENEYEVLYKPLAGVRSKLOENNYM 1306
Qy 1307 FNVVYKDLNSRPNKRENFKNVLESDLIPYKDLTSSNVKDPYKELNKKERDQFLSSYN 1366
Db 1354 FNLNIDNLSRKLKRYKFLDVLDESQFQKHISSENEYIEDSFKLNSDOKNTLKS 1413
Qy 1367 YIKOSIDIDINPANDVGLGYKILSKYSKSDLSIKKYI-----NDK 1407
Db 1414 YIKESVENDIKFAQEGISYERKYLAKYKDDLESIKKVIKEKEPPSPPTPPAKTDE 1473
Qy 1408 OGENEKYLPFLNLTETKTYNDKIDLEFVILHEAKVLWYTEKSNVEKIKELNYLKTIO 1467
Db 1474 OKRESKFLPELNETIETLYNNLVNKITDYLINLAKINOCNVEKDEAHKIKLSLKA 1533
Qy 1468 DKLADFKNNNFVGIADISTDYNNHNLTKFLSTGQWFEENLAKTVLSMLDGNLQ 1527
Db 1534 DKIDLFKNTNDFEALIKLINDYTKKMDLGLSTGLV-QNPPNTITSKILGKFKQDM 1592
Qy 1528 SOHCYKQKQCPQNSGCFHNDERECKLLNWKQBGKCVENPNTCENNNGCDADAKC 1587
Db 1593 SOHCYKQKQCPQNSGCFHNDERECKLLNWKQBGKCVENPNTCENNNGCDADAKC 1592
Qy 1588 TEEDSGSNKRTCECTKPDSPYPLFDGIFCSCSSNFGISFLILMLILYSFI 1639
Db 1653 TEEDSGSNKRTCECTKPDSPYPLFDGIFCSCSSNFGISFLILMLILYSFI 1704

RESULT 2
Q9TZT5 PRELIMINARY; PRT: 1694 AA.
AC Q9TZT5.
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE MEZOZOITE SURFACE PROTEIN 1 PRECURSOR.
GN MSPL.
OS Plasmidium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmidium.
OX NCBI.TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HN1;
RA Jiang G., Liu R.Z., Daubenberger C.A., Pluschke G.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF062348; AAC72884.1; -.
DR InterPro; IPR000561; -.
DR Pfam; PF00008; EGF_1.
KW Merozoite.
SQ
SEQUENCE 1694 AA; 192794 MW; 84CFCD0E709F5673B CRC64;

Query Match 61.3%; Score 5162.5; DB 5; Length 1694;
Best Local Similarity 61.1%; Pred. No. 6.1e-183;
Matches 1054; Conservative 226; Mismatches 328; Indels 117; Gaps 27;

```

QY 241 YIKKAKTIEININLEIESKKTIDKNNKATKEEKKKLYOAYOLSTYNNKOLEAHNLIS 300
 DB 235 YIKKAKTIEININLEIESKKTIDKNNKADNEEKKKLYOAYOLSTYNNKOLEAHNLIS 294
 QY 301 VLEKRIOTLKKNEIKELLOKINEIK-NPPRANGNTPNLLD-KNKKIEHEKEIKIA 358
 DB 295 VLEKRIOTLKKNEIKELLOKINEIK-NPPRANGNTPNLLD-KNKKIEHEKEIKIA 354
 QY 359 KTIKFNIDSLFTDPLELEYILREKNKIDISAKVE--TKESTERN-EYRNGVTPYLSYND 415
 DB 355 KTIKFNIDSLFTDPLELEYILREKNKIDISAKVE--TKESTERN-EYRNGVTPYLSYND 414
 QY 416 INNAL--NELANGDILNPDYTKEPSKNTYTON-ERKATINEIKERIKIEKKI---- 467
 DB 415 INNAL--NELANGDILNPDYTKEPSKNTYTON-ERKATINEIKERIKIEKKI---- 467
 QY 468 ESOEKSYEDSKSLNDITKEVEKLLNELYOSKFNNNIDLTNFEKMGKRSYKVEKLTJH 527
 DB 474 ESOEKSYEDSKSLNDITKEVEKLLNELYOSKFNNNIDLTNFEKMGKRSYKVEKLTJH 529
 QY 528 NTFASYSKHNLEKLTALKYMEDYSLRNIVREKELKYNNLSKIEENETETVENIK 587
 DB 530 NTFASYSKHNLEKLTALKYMEDYSLRNIVREKELKYNNLSKIEENETETVENIK 589
 QY 588 DEOLOFEK--ITKDEKRPDEKLEVSIVYOVYKVLANKIDELKKTOLILKNEWLK 644
 DB 590 DEOLOFEK--ITKDEKRPDEKLEVSIVYOVYKVLANKIDELKKTOLILKNEWLK 645
 QY 645 HNIVPVSYSKHNLEKLTALKYMEDYSLRNIVREKELKYNNLSKIEENETETVENIK 704
 DB 646 HNIVPVSYSKHNLEKLTALKYMEDYSLRNIVREKELKYNNLSKIEENETETVENIK 704
 QY 705 TEGITQOATTKPGQOAGSALSGDSVOAQOEOKQOAPVPEVPEAKAOVPTPAPV-- 762
 DB 697 TEGITQOATTKPGQOAGSALSGDSVOAQOEOKQOAPVPEVPEAKAOVPTPAPV-- 762
 QY 763 TEGITQOATTKPGQOAGSALSGDSVOAQOEOKQOAPVPEVPEAKAOVPTPAPV-- 762
 DB 754 TEGITQOATTKPGQOAGSALSGDSVOAQOEOKQOAPVPEVPEAKAOVPTPAPV-- 762
 QY 814 ESKLSQDPLDLFNIONNIPVMSMPSDLNLSQLEMEIEREKEMCNLYKLKDNKIK 873
 DB 814 ESKLSQDPLDLFNIONNIPVMSMPSDLNLSQLEMEIEREKEMCNLYKLKDNKIK 873
 QY 874 NLEBAKV-----STSVKTLSSSS-----MQLSLTPDOKPREVAND---D 912
 DB 874 NLEBAKV-----STSVKTLSSSS-----MQLSLTPDOKPREVAND---D 912
 QY 913 TSHS-----TNLNNSLKLFENILSGKKNKIYQEL-IGQKSENEYEKILKOSDFYNSF 967
 DB 934 TSHS-----TNLNNSLKLFENILSGKKNKIYQEL-IGQKSENEYEKILKOSDFYNSF 967
 QY 968 TNPYKSKADNINSINDESKRRKLEEDIKLTKTOLSFDLNKKYKLEKLFEDKKKTYGK 1027
 DB 994 TNPYKSKADNINSINDESKRRKLEEDIKLTKTOLSFDLNKKYKLEKLFEDKKKTYGK 1027
 QY 1028 YKMOIKKLTLEKLEKSLNSLNNPKHVLONFVFPNKKKAEALAEENTLENTKILKH 1087
 DB 1053 YKMOIKKLTLEKLEKSLNSLNNPKHVLONFVFPNKKKAEALAEENTLENTKILKH 1087
 QY 1088 YKGLVYKNGSSPLKTLSESIOTEDNYASLENPKYVSKLEGLKXDLNLEKKKSLYS 1147
 DB 1113 YKGLVYKNGSSPLKTLSESIOTEDNYASLENPKYVSKLEGLKXDLNLEKKKSLYS 1147
 QY 1148 YKGLVYKNGSSPLKTLSESIOTEDNYASLENPKYVSKLEGLKXDLNLEKKKSLYS 1147
 DB 1173 YKGLVYKNGSSPLKTLSESIOTEDNYASLENPKYVSKLEGLKXDLNLEKKKSLYS 1147
 QY 1202 -----SGSDTLQOQPKRPASTHVGAESNTJTSQNVDEVDVIVIPFGSEEDY 1253
 DB 1232 -----SGSDTLQOQPKRPASTHVGAESNTJTSQNVDEVDVIVIPFGSEEDY 1253

QY 1254 DDGLQVNTGEAVTSPVIDNINILSKITENEYEVLYLKPLAGVRSLLKKOLENNVFNENVKD 1313
 DB 1292 DDGLQVNTGEAVTSPVIDNINILSKITENEYEVLYLKPLAGVRSLLKKOLENNVFNENVKD 1313
 QY 1314 ILNRFNKNRNFKNFVLESDLPYKDLTSSNVVADKPYFLNKKERKDFLSSYNYIKSID 1373
 DB 1351 ILNRFNKNRNFKNFVLESDLPYKDLTSSNVVADKPYFLNKKERKDFLSSYNYIKSID 1373
 QY 1374 TDINFANDVIGYKILSEKYSIDSLIKKYI-----NDKOGENEKY 1414
 DB 1411 TDINFANDVIGYKILSEKYSIDSLIKKYI-----NDKOGENEKY 1414
 QY 1415 LPLNTIETLYKTYNDKIDLEFVHLEKVLNVTYKESVNEKIKELNWTYKTOXKLDK 1474
 DB 1471 LPLNTIETLYKTYNDKIDLEFVHLEKVLNVTYKESVNEKIKELNWTYKTOXKLDK 1474
 QY 1475 KNNFVGIADLSTDYNNHNLTKFLSTGMVFNELAKTYVLSMLDGNLQGMINISQHOVC 1534
 DB 1531 KNNFVGIADLSTDYNNHNLTKFLSTGMVFNELAKTYVLSMLDGNLQGMINISQHOVC 1534
 QY 1535 KOPQNSGCFRHLDERRECKCLNLYKOGDKCVENPNTENNNGCDADAKTEEDSGS 1594
 DB 1590 KOPQNSGCFRHLDERRECKCLNLYKOGDKCVENPNTENNNGCDADAKTEEDSGS 1594
 QY 1595 NGKRTCECTKPDSPYPLFDGIFGSSNPLGTSFLLMLLYSPL 1639
 DB 1650 NGKRTCECTKPDSPYPLFDGIFGSSNPLGTSFLLMLLYSPL 1639
 RESULT 3
 Q9NHX1 PRELIMINARY: PRT: 1694 AA.
 AC Q9NHX1: 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE MAJOR MEROZOITE SURFACE ANTIGEN.
 GN GP195.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FCCL/HN.
 RA Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M.,
 RT "Molecular cloning and sequence analysis of major merozoite surface
 antigen (gp195) gene of Plasmodium falciparum isolate FCCL/HN."
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF218248; AAF27526.1;
 DR InterPro: IPR000561;
 DR InterPro: IPR003247;
 DR Pfam: PF00008; EGF 1;
 DR ProDom: PD001527; -; 1.
 KW Merozoite.
 SQ SEQUENCE 1694 AA: 192766 MW: B51634A49E0F6728 CRC64:
 Query Match 61.3%; Score 5162.5; DB 5; Length 1694;
 Best local similarity 61.1%; Pred. No. 6.1e-183;
 Matches 1054; Conservative 226; Mismatches 328; Indels 117; Gaps 27;


```

OY 181 LNEYEDLLRAKINDVANCANDYCOIPENLIRANELDLVLEKLVGKRPDLNIDKNDKMD 240
DB 175 LNEYEDLLRAKINDVANCANDYCOIPENLIRANELDLVLEKLVGKRPDLNIDKNDKMD 234
OY 241 YIKNNKTIININILIEESKKTIDKNNKREKKEKLLYQAOYOLSTYNNOLEAHNLIS 300
DB 235 YIKNNKTIININILIEESKKTIDKNNKREKKEKLLYQAOYOLSTYNNOLEAHNLIS 294
OY 301 VLEKRIOTLAKNENIKELDLKINEIK-NPPRANGNTPFLTD-KNKKIEHEKEIKEIA 358
DB 295 VLEKRIOTLAKNENIKELDLKINEIK-NPPRANGNTPFLTD-KNKKIEHEKEIKEIA 354
OY 359 KTIKENIDSLFTDPLEEYVLRKKNIDISAKVE--TKRSTERN-EYPNGVTPYLSYND 415
DB 355 KTIKENIDSLFTDPLEEYVLRKKNIDISAKVE--TKRSTERN-EYPNGVTPYLSYND 414
OY 416 INNAL---NELNSFGDILNPDYTKERSKNITPNN-ERKFINIKERKIKIEKKI---- 467
DB 415 INNAL---NELNSFGDILNPDYTKERSKNITPNN-ERKFINIKERKIKIEKKI---- 467
OY 468 ESDKSSYEDRSKSLNDITKEVEKLLNEIYDSKFNNNIDLTNFEKMGKRSYKVEKLTTH 527
DB 474 ESDKSSYEDRSKSLNDITKEVEKLLNEIYDSKFNNNIDLTNFEKMGKRSYKVEKLTTH 529
OY 528 NFEASYENSKHNLEKTLKALYMEDYSLRNIYVEKELKTYKNLISKIENIEIETLVENIKK 587
DB 530 NFEASYENSKHNLEKTLKALYMEDYSLRNIYVEKELKTYKNLISKIENIEIETLVENIKK 589
OY 588 DEBOLEPKR---ITKDNKRDKEKLEVSOLYKVOYVLLNKLDELKKTOLILKYNELK 644
DB 590 DEBOLEPKR---ITKDNKRDKEKLEVSOLYKVOYVLLNKLDELKKTOLILKYNELK 645
OY 645 HNHLVPSYKQENKOEYVYLIVLKEIDKLVMPKVESILNBEKKNIKTEGSDNSEPS 704
DB 646 HNHLVPSYKQENKOEYVYLIVLKEIDKLVMPKVESILNBEKKNIKTEGSDNSEPS 704
OY 705 TEGELTGATTKRGOQAGSALESDSYQAOEQKQAPVPVPVPEAKAOPVPAPAV-- 762
DB 697 TEGELTGATTKRGOQAGSALESDSYQAOEQKQAPVPVPVPEAKAOPVPAPAV-- 753
OY 763 -----NNKTEVNSKLDYLEKLYEFLNTSYCHKTYIIVSHSTNMEKLLKOKYITKEE 813
DB 754 VENSIEHKSMDNSOALTKYVLLKLEFLKYSYCHKTYIIVSHSTNMEKLLKOKYITKEE 813
OY 814 ESKLSSCDPLDLFNIONNIPVMSMEDSLNNSLSOLFMEIYERKEMVCNLYKLDNDKIK 873
DB 814 KNEKSGCDPLDLFNIONNIPVMSMEDSLNNSLSOLFMEIYERKEMVCNLYKLDNDKIK 873
OY 874 NLEBAKV-----STSVKTLSSSS-----MQLSLTPQDKREVSAND--D 912
DB 874 KLEEQKQITGSTSTSPGNTVTNTAQSATHNSOQNSASSSTNTONGVAVSSGPAVVE 933
OY 913 TSHS-----TNLNSIKLFENILSIGKKNKIYQEL-IGOKSESENYEKLILDSOTFYNESF 967
DB 934 ESHDPLTVLISIDMLGIYSLNLDGKTKVNPPLTJSTTEMEKFEYENILKNNDTYFNDI 993
OY 968 TNFVSKASDINSLENDSEKREKLEEDINKLKTLOLSFDLYNNKYKLELRFDKKATVKG 1027
DB 994 KOVVKSMKYITGLT-ETQKNALNDEIKKLKDLQLSFDLYNNKYKLELRFDKKATVKG 1052
OY 1028 YKNOIKKLLILKQLESKSLNLPKPHVLONFVSFENKKEAEIETENTLENTKILKH 1087
DB 1053 DKMOIKKLLILKQLESKSLNLPKPHVLONFVSFENKKEAEIETENTLENTKILKH 1112
OY 1088 YKGLVYNGESSPLKTLSESTOTEDBNVASLENFKYLSLEKGLKNDNLLEKKKLSYLS 1147
DB 1113 YKGLVYNGESSPLKTLSESTOTEDBNVASLENFKYLSLEKGLKNDNLLEKKKLSYLS 1172
OY 1148 SGLHHLIAELKEVIKKNKNTYGNSPENNTDYNNALEKFKPEGTGVATVASE----- 1201
DB 1173 SGLHHLIAELKEVIKKNKNTYGNSPENNTDYNNALEKFKPEGTGVATVASE----- 1231
OY 1202 -----SGSDTLEOSQPKKPASTHVAGESNTITTSQNVDEVDVYIIVIFGESEEDY 1253

```

```

DB 1232 PSLPVSRRSSGSGSTKEBETQIPGSGSLTTELQVQSONDEEDDSLVLP-IFGESEEDND 1291
OY 1254 DDGQVVTGEAVTPSVADNITLSKIENIEYVLYKPLAGVYRSLLKQLENNVMTNVAVKO 1313
DB 1292 EYLDQVVTGEAVT-NDNLISGENEYDVYLYKPLAGVYRSLLKQLENNVMTNVAVKO 1350
OY 1314 ILNSRFKRENFKNVLESDLDIPYKDLTSSNVVVKDPKFLNKEKRDKFLSSYNTKSID 1373
DB 1351 ILNSRLKRRKFLDVLSESDLMQFHKHISNEIIEDSRKLNSBQKNLILSKYTIKRSVE 1410
OY 1374 TDINFANDVLGYKYLSEKYSKSDISIKKYI-----NDKQGENEKY 1414
DB 1411 NDIFAEQGISYEVKYLAKYKDDLESIKKYIKEKEEPSSPPTTTPSPAKTDEQKRESKE 1470
OY 1415 LPLNLNITLYKYVNDKIDLFVHLEAKVNLTYTEKSNVEYKIKELNYLKTIDQKLDFK 1474
DB 1471 LPLNLNITLYKYVNDKIDLFVHLEAKVNLTYTEKSNVEYKIKELNYLKTIDQKLDFK 1530
OY 1475 KNNFVGIADSLDTPDYNHNNILTKFLSTGAVFENLAKTVLSNLDGNLGMLNISQHCYK 1534
DB 1531 NTNDFEAIKLLINDPTKMDLKGILSTGLV-QNFPNTIISKLEKQDMNLINSQHCYK 1589
OY 1535 KQCPONSCEFRHLDEREKCCLLNYKQEGDKCVENPPTCNENNGGCDADAKTEEDSGS 1594
DB 1590 KQCPENSCEFRHLDEREKCCLLNYKQEGDKCVENPPTCNENNGGCDADAKTEEDSGS 1649
OY 1595 NGKRTTECTKPDSPYLFEDGIFGSSNFIPLILMLILYSFI 1639
DB 1650 NGKRTTECTKPDSPYLFEDGIFGSSNFIPLILMLILYSFI 1694

```

```

RESULT 4
ID 025922 PRELIMINARY; PRT: 1720 AA.
AC 025922;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-MAR-2001 (Tremblrel. 16, last annotation update)
DE PRECURSOR OF THE MAJOR MEZOZOITE SURFACE ANTIGENS.
OS Plasmodium falciparum (isolate NF54).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5843;
RN [1]
RP SEQUENCE OF 1069-1720 FROM N.A.
RC STRAIN=NF54;
RA Tolle R., Bujard H., Cooper J.A.;
RL Exp. Parasitol. 0:0-0(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NF54;
RA Tolle R.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=NF54;
RX MEDLINE=96123395; PubMed=8577332;
RA Pan W., Tolle R., Bujard H.;
RT "A direct and rapid sequencing strategy for the Plasmodium falciparum
antigen gene gp190/msal."
RL Mol. Biochem. Parasitol. 73:241-244(1995).
DR EMBL: 235327; CAA84556.1; -
DR InterPro: IPR000561; -
DR Pfam: PF00008; EGF_1.
KW Signal; Merozoite.
FT SIGNAL 1 19 POTENTIAL.
SQ SEQUENCE 1720 AA; 195725 MW; 717B2FD1E637A8A3 CRC64;

```

Query Match 60.6%; Score 5107.5; DB 5; Length 1720;
 Best Local Similarity 60.2%; Pred. No. 6; 6e-181;
 Matches 1052; Conservative 216; Mismatches 343; Indels 137; Gaps 29;

```

QY 1 MKIIFELCSFLFIINOCVTHESYOELVKKLEALDVAITGYSLFQKEMVLANE--GT 57
D 1 MKIIFELCSFLFIINOCVTHESYOELVKKLEALDVAITGYSLFQKEMVLANEELTT 60
QY 58 SGTAVTSTPGSKSVASGSGSVASGSGSVASGSGSVASGSGCN--SRFTN-- 112
D 61 KGASQSGASQSG--ASQSGASQSGASQSGASQSGSTGSPSGTSPSSKSNLTP 118
QY 113 -----PSDNDSDAKSYADLKHVRNVLITIKELKPOLFEDLNNHMLTCDNHG 163
D 119 RSNMSSGASPADASDQSAKSYADLKHVRNVLITIKELKPOLFEDLNNHMLTCDNHG 178
QY 164 FKYLIDGEEINELLYKLFNFDDLRAKLVANDVCANDYCOIFPNLKRANELDVLKLVFG 223
D 179 FKYLIDGEEINELLYKLFNFDDLRAKLVANDVCANDYCOIFPNLKRANELDVLKLVFG 238
QY 224 YRKLPLNDIKDQVGMEDYIKKNNKTTIENINELIEBSKKTIDKNNKATKEEKKLYQAY 283
D 239 YRKLPLNDIKDQVGMEDYIKKNNKTTIENINELIEBSKKTIDQNNKADNEGGKKLYQAY 298
QY 284 DLSIYNKQLEBAHNLISYLEKRIDTKKNNIKELDKINIKPNPPANSNGTPTLTDK 343
D 299 DLSIYNKQLEBAHNLISYLEKRIDTKKNNIKELDKINIKPNPPANSNGTPTLTDK 358
QY 344 NKIEHEKEIKELIAKTIKFNIDSLFTDPLELEYLRKNNIDISAKVE--TKESTEPN 401
D 359 NKIEHEKEIKELIAKTIKFNIDSLFTDPLELEYLRKNNKVDTPKSQOPTKVOQIPK 418
QY 402 -EYNGVYPLSYDINNAL--NELNSFGDLNPDYTKER--SKNITYDN-ERKKFTNE 455
D 419 VPYNGVYPLSYDINNALNDKNSYGDLMNP--HTKEKINIKITJDKERKIFINN 476
QY 456 IKKIKIKIKKI-----ESDKSYEDRSKSLNDITKEVEKLNLEIYDSKNNIDLTNEK 511
D 477 IKKIKIDLEBKININTKEONKRLDEYKS---KDYELLEKFEKMFNNNPDVDVKK 532
QY 512 MMGRYSYKVEKLTHTNHFASYENSKHNLKELTKALYMEDYSLNIYVEKELKYYKMLI 571
D 533 IFSARYTYNVEKQRYNNKFSSSNSVYVQKTKALSLYEDYSLRKGISEKDFNNITYLK 592
QY 572 SKINEIETLVENIKKDEQLFEKK--ITKDEKPDKILEVSDIVKVOQVYLLMKI 628
D 593 TGLBADIKKILEEIKSEKNIKLEKNFKGLTHSANGS---LEVSDIVKVOQVYLLMKI 648
QY 629 DELKKTQILKNVLEKLNHNHPNSYKQENKQEPYLLIYKLEIDKLYPMKVESLINEE 688
D 649 EDLRRKIELFKNAOLKOSIHVPNIYKPKQNKREPYLLIYKLEKREVDKLEFIRKXMKLKE 708
QY 689 KKNIKTEGQSDNSEPSTGECITGOATTPKQOAGSALBGDSVQAOQOKOAO-----P 742
D 709 QAVLSS-----ITQPLVAASETTEDGSHTHITLSQSETEVTEETEETEVGHTT 759
QY 743 PVVPVPEAKAQPPTPAVY-----NNKTENVSKLDYLELEKLEFINTSYICHKY 791
D 760 TVTITLPTQ---PSPREKVUVENSIHEKSDNSQALTKVYLYKLEDEPLTKSVICHKY 816
QY 792 ILVSHSTNENKILQYKITKEESKLSLCPDLDFNIONNIPWYSKFPDLSNLSOLF 851
D 817 ILVSNSSMDOKLLEVYNTPEEENELKSCDPLDFNIONNIPWYSKYDSDNNDLQHLF 876
QY 852 MEIEREKWVCULYKLDKDKLKNLLEEKVY-----STSVKTLSSSS----- 893
D 877 FELYQKEKITIYLLHKLKEENHKLLEBOKQITGSSSTSPGNTTVNVAQSTHNSQOQ 936
QY 894 MOPILSTPQDKEVVSAND--DTSHS---TNLNSKLKLFENILSKKNNIYOEL-IGQ 945
D 937 SNASTTNQNGVAVSSGPAVEESHDPPLTVLISNDLKGIYSLNLGKKTVPNPLTST 996
QY 946 KSSNFYEKILKSDTFNENEFNVVSKADDINSLNDSEKRRKLEEDINKLKTLOLSF 1005
D 997 TEMEFYEENIKNDYFNNDIKQFVSNKSVITGLT-ETQKNLNLDEIKKLKTLOLSF 1055
QY 1006 DLYNKKYKTLERLDDKKTGVKYMQIKKLTLKQLESKLNLSLNNPNNHVLQNSVFENK 1065

```

```

D 1056 DLYNKKYKTLERLDFENKKEKLGODKMQIKKTLTKQLESKLNLSLNNPNNHVLQNSVFENK 1115
QY 1066 KKEAIAETENTLNTKILKHKYKLVYNGESSPLKTLSEESIQTODNVALENFYL 1125
D 1116 KKEAIAETENTLNTKILKHKYKLVYNGESSPLKTLSEESIQTODNVALENFYL 1175
QY 1126 SKLEKTLKDNILNEKKKISLYSSGLHNLIAELKEVYKKNYTGNSPNNNDVNNALSY 1185
D 1176 SKIDKLDNMLHGLKKKLSFLSSGLHNLITELKEVYKKNYTGNSPNNNDVNNALSY 1235
QY 1186 KFLPEGTVAIVVSE-----SGSDPLEOSQPKRPASTVGAESNTITTSQ 1231
D 1236 ENFLPE-AKVYTVVTPPPDPVTPSPLSVRSVSSGSTKEETQPIPTSGSLINELQOVQLO 1294
QY 1232 NVDEEDVIVITPIFGESEEDYDGLGVYTGAPVTPSYIDNLSIKIENEVEYLYKPLAG 1291
D 1295 NYDEEDDSLVLVLPITFGESEEDNDEYLDQVVTGSAISVT--MDNILLSGFENEYDVIYKPLAG 1353
QY 1292 VYRSLKQLENNVMTFNVNDILNSRFNKNREPNKVEESDILIPYKDLTSSVYVVKDPPK 1351
D 1354 VYRSLKQLENNVMTFNVNDILNSRFNKNREPNKVEESDILIPYKDLTSSVYVVKDPPK 1413
QY 1352 FLNKEKRDKFLSSYNYIKOSIDTIDINFANDVLYGYKILISEYKSDLSIKRYI----- 1404
D 1414 LNSQKQWTLKSYKIKESYENDIKFAQEGISYEKVLAKYKODLESIKKVIKEKEKF 1473
QY 1405 -----NDKQENKYLEFPLNNITELKYTVNDKIDLYIHLKAVLNTYKS 1451
D 1474 PSSPPTTPSPAKTDEQKESKFLPFLTNIELVNNLKNKIDDYDILNKAINDCNVEKD 1533
QY 1452 NVEVYKELNLYKTIDKLADEPKKNNFVGLADISTDVNNHNLKFLSTGVFENLKT 1511
D 1534 EAHVATITLSDKALDXXIDIDLEKPNYDEALIKKLINDOTKKOMLKLSTGLV-QNEPNT 1592
QY 1512 VLSNLDGNLQGLMNIISOHCYKQCPONSGCFRHLDERRECKCLLNYKQEGDKCVENPN 1571
D 1593 IISKIEKFPQDMLNISQHCQKQCPENSGCFRHLDERRECKCLLNYKQEGDKCVENPN 1652
QY 1572 PTCNENNGCCDADAKCTEEDSGSGKKTITCECTKPDSPLEDFGIFCSSNPLGSLFLLIL 1631
D 1653 PTCNENNGCCDADAKCTEEDSGSGKKTITCECTKPDSPLEDFGIFCSSNPLGSLFLLIL 1712
QY 1632 MLILXSF 1639
D 1713 MLILXSF 1720

```

RESULT 5

Q26194 PRELIMINARY; PRT: 1751 AA.

AC Q26194; 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)

DE (PV200).

GN PV200.

OS Plasmodium vivax.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_Taxid=5855;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SAU-1.

RX MEDLINE=92158013; PubMed=1371329;

RA Gibson H.L., Tucker J.E., Kaslow D.C., Kretzli A., Collins W.E., Kiefer M.C., Bachurst I.C., Barr P.J.;

RT "Structure and expression of the gene for Pv200, a major blood-stage surface antigen of Plasmodium vivax";

RL Mol. Biochem. Parasitol. 50:325-334(1992).

DR EMBL; M75674; AAA29735.1; -

DR HSSP; P05412; 1JUN.

KW Signal.

FT SIGNAL 1 19 POTENTIAL.

FT CHAIN 20 1751 POTENTIAL.
SQ SEQUENCE 1751 AA: 196120 MW: 41629883F29D8A6 CRC64:

Query Match 36.7%; Score 3093.5; DB 5; Length 1751;
Best Local Similarity 38.2%; Pred. No. 1,4e-106;
Matches 704; Conservative 316; Mismatches 517; Indels 307; Gaps 41;

QY 1 MKTIFELCSFLFIINOCYTHESYOELVKKLELDAVLGYSLPQEKRYVNEGSGT 60
DB 1 MKALLFSEFEEFYTKQCCET-ESYKQLVAKLDKLEALVDGVEFLFKKKL----- 50
QY 61 AVTSTPGSGSVASGSGSVASGSGSVASGSGSVASGSGSVASGSGSRTRPNSDS 120
DB 51 -----GENDIKVETNANNNNN 68
QY 121 DAKSYADLKHRVNYLLTTELKYPQLEFDLTNMLTLCODIHGKYLLIDGYEELNELLYK 180
DB 69 NQSVV--LTSKIRNLFSLKLELOIPGHITDLHLIRELAVEPNGIKYLVESYEENQIMHV 126
QY 181 LNFYFDLLRAKLVNCANDYQOIFPNLKIRANELDVLKLLFGRKPLDNKDWKMD 240
DB 127 INFHYDLRLAKHDMCAHDYCKIPKHLKISDKELMDLKKVLRKPLDNKIDIGKLET 186
QY 241 YIKNKKTIENINEL-TEESKK----- 261
DB 187 FITNKKTIKINISDLIIAENKKRSGHPTTTNGAGTQPANSGIAASSETQISGSSNSG 246
QY 262 -----TIDKNKATKEEKKKLYQAOYDLSYKQ 291
DB 247 SSSSTGSSNSGSSSTGSGTGTGQSPPAADASTNANEAKKIYQAVYNTIFYTNQ 306
QY 292 LEEAHNLSVLEKRIIDLTKKNEIKELDKINELKNPPA---NSGNPNLLDKNKIE 348
DB 307 LQEQKLAIVLEKRVKVKLEKHKDILVLEQVAKKEKLPSPYPTNTLTNHNKESKIA 366
QY 349 EHEKEIKETIAKTIFNIDSLTDPLELEYLREKNKNI-DISAKVETKESTEPN----- 402
DB 367 ELEKIEIAKTAVNFDLDGLFTDAELEVLREKAKMAGLIIPESTRKSACTPGCTVPTL 426
QY 403 ---YPCGVYTYLSTNDINNALNELNS---FGDLNPFDPYTEPSKNI-YTNERKKFTNE 455
DB 427 KETYPHGISYLAENSIYELLEKIGSETFGDLQNP--DDGQPKKGLIINETKRELEK 485
QY 456 IKEKIKIEKKIESDEKSYEDRSKSLNDITKEYEKILNEIYDSFPNNNIDLTPEKMMGX 515
DB 486 IMANKIKIEEDKLPULKKEEYKVEKAVNEKPAFNFHFEARDNTLVENKEDDEPKK 545
QY 516 RYKSVKELTHTNFTASYENSKHNLKLTAKLYMEDYSLRNIYVEKLYKXNLISKE 575
DB 546 REAVMEKKKLES--CSYEQNSNLINKLKQLYLEADVLRKDIADDEIKHFSPEMKLK 603
QY 576 NEIETLVENIKDEBOLFEEKITKDEKRPDKILEVSDIYVVOYKVLMMKIDELAKTQ 635
DB 604 SEIDVLAQOEIKKRNENKL-----TIENK-----FDFSGVVELQYOKVLIKKIEALKNVQ 652
QY 636 LILKNVELKHNIHPNSYKQENKQEPYLIYLKKEIDKLKFMKPVESLIEEKK----- 690
DB 653 NLLNNAKVKDLYIPKYVKTSEKPEPYLMLVKREIDKLKFIKIESMATEKKKPIYA 712
QY 691 --NRTKGOS--DNSEPESTEG-----EITGOATTKRG-----QAQSA--LE 726
DB 713 AADIVANGQSLRGASSETGTGNTVNAQTAVVQPHQVNAVATYOGTGHQAQGEAEATQ 772
QY 727 GDSYQAOAQEQKQA-----QPPVPPV-PP-AKAQVTPPAPVNNKTEVNSK 771
DB 773 TNSYQAOAVQCTPAGAGGQVASTOTISQAPAPTAQSPAPAPAPSTPAAVAPAPATMSK 832
QY 772 LDYELKYEELINTSYGKRYLLVSHSTNEKILKQYKITEESKL--SSCDPLDLLNI 829
DB 833 LEYELKLDLFLSAKACHKHLFVTSNITKHELDDOYKLNADENKINTKCDDELDFNV 892
QY 830 QNNIPVMTSMFDSLNNSLSQLFMEIYEKEMVCLYKLDND-KIKNLEAKK-----VS 883

DB 893 QNNIPVMTSMFDSLNNSLSQLFMEIYEKEMVCLYKLDND-KIKNLEAKK-----VS 883
QY 884 TSVKTLSSSKQPLSLT-PODKPEVSANDTSHSTNLSNLKLEPNLISLQKKNITQEL 942
DB 953 SAAPSCQAGTTPVTTAPVTTTTPPSPOISVTSPPPOAEN--RRVGSNEEKPE- 1010
QY 943 IGQKSENFYEKILKDSDFTFNEFSNTEKVSADISLNDSEKRRKLEEDINKLTKIQ 1002
DB 1011 ADTAQVEKFEYKHLSDDK-YNDYFOKFLSEQKDEITKM-DEYWKVLDGAEIELEKKLQ 1068
QY 1003 LSPDLNKKYKLELERLEDDKKTGYKRYKQIKKLTLEQLESKUNSLNPKVHLQNSVF 1062
DB 1069 VSLDHGYKYKLELERLKKKIKNSKQIKKLTSLKNKLEBRONLNNPVSYLKNTAF 1128
QY 1063 FNKKKEADIAETENTLNTKILKHGKLVKYNGESSPLKTISEESIQEDNVALENF 1122
DB 1129 FNKKRETEKEVENTLKTNTELLKYYARAKYVIGEPPLKTISEESMOKEDNLTNEKF 1188
QY 1123 KVLKLEKLEKDNINLEKKKLSYLSGLHLIAELKEVYKKNKNTGNSPSENNTDVNNA 1182
DB 1189 RYLSRLEGRGLKNIELKENISTYSSGLHHVLTLEKELIKNNKYSGNDHTKNIAVREAL 1248
QY 1183 ESKKFLPEGTAVVVSSESGDTLEQS-----QPKPA- 1216
DB 1249 QAYQELIPK-----YTTQEGASTTAATLPTVPSAVPGCLPGAGVGAAGLTPPPAG 1302
QY 1217 ---STHYGA-----ESNTITTSQNVDEVDYITVIFGSEEDYDLOGVYNGEA----- 1264
DB 1303 SVPATGPAAAGSTEENVAKAQYADYDVKVIALPFGNDDGEE-DOVTTGAESEA 1361
QY 1265 ---VTPSYIDNLSKIENEYEVLYKPLAGYVRSKQLENNVMTFVNVNVDILNSRFNK 1321
DB 1362 PEILVPAGI-----SDYVVYVLTPLAGMYKTIKKQLEHNVAFPNTITDMLDSLKK 1413
QY 1322 RENRKNVLESLLPYKDLTSSNVYVQDPYKFLNKKKDKFLSSYNYIKDSITDIDINPAND 1381
DB 1414 RNYFLEVYLSNDLNFKSSSGEYIYKDPYKLLDEKKKKLIGSYKIGASIDMDLAPAND 1473
QY 1382 VLGYYKILSEKYSKDSLOSIK---KYIND-----KQGEN 1411
DB 1474 GVTYVNNKKGELVYKTHLDGVKYEIKRVEDDIKKODELKKLCNVANSODSKNEFIATAKREL 1533
QY 1412 EKYLPFLNNIETLYKTYVNDKIDLEFVILHLEAVLVNTYKESNVEYKIRELNYLKTIOKLA 1471
DB 1534 EKYLPFLNSLDKXESLVSKYNTYTDNLKVIYNNQOLEKKEAEITVYKLOYNMDEKLE 1593
QY 1472 DFKKNNNEVGIADISTDYNHNNLLTKFLSTGVFENLAKTYLSMLDGNLOGMLNISQHQ 1531
DB 1594 EYKR-----SEKKNVEKSSGLLEKLMKSKLIRENESKEITLSQILNVOTDLTMSSEHT 1646
QY 1532 CVKCKQPOKSGCFHLDREKCKCLNLYKQEGDQVEMNPCTGNNNGCCADAKTEED 1591
DB 1647 CIDTNVPDNAACVRLDDETEWRCLLTFKEGKCVAPASNTCTCDNNGGCAPEADCKMTD 1706
QY 1592 SGNNGKKTICECTPDSPYLPDGFICSSSNFLGISFLILMLIL 1635
DB 1707 S-----NKIVCKCTKEGSEPLFEGVFCSSSFLSLFLMLLFL 1746

RESULT 6
025961 PRELIMINARY; PRT; 1087 AA.
AC 025961;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE MAJOR SURFACE ANTIGEN (FRAGMENT).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.

Db	931	NDNC--ASNHNHNSYSKSPGNI--SCAKHSTPQAEENQVGVGENSEKPEADPAQAEKPEYDKHL	988
Qy	957	KDSDFPNESFTNFVWVSKADQINSLNDESKRKKLEEDINKLKTQLQSFGLYNKTKKLE	1016
Db	989	SQIDK-VNDYEFKKFLESKEELIKM--DPTKMAWALGEIEELKKKLOVSLDHGKTKKLE	1046
Qy	1017	RLEPKKTKVAKYKMOAKIKLLKLEKOLESKLNSLNPNKHLYQNFSEVFNKKKEAEIAETEN	1076
Db	1047	RFLKKKKKINSKDOIKKLLSLKNKLEBRONLNNPTSLKNTYTPAFNNKRTTEKKEVEN	1106
Qy	1077	TLENFKILKHKYGLVYVNGESSPLKTSIESIQTEDNYVASLENFKYLSKLEGLKDNL	1136
Db	1107	TLKTEITLLKYKAKAKAYYIGEPFPLKTSLESMQEDNYLMLEKFRCSADMK--EIKDQT	1165
Qy	1137	NLEKKLSIYSSGGLHLLAEKLVKNNKNTYGSBENNTDVNNALSEKTKLP-----E	1191
Db	1166	ELEKSNISYSSGLLHLYDAEELINDKRYSCGDHAKNAEKKKLAQVYQELIPVYSQE	1225
Qy	1192	GTDAVAVV-----SESGSDTLEQSPKKPASTHY--	1220
Db	1226	STSAVYVPCGAVVPCVPTAAAGSGAGVPPAAAGSASCAVPPAGCPSPATGVVP	1265
Qy	1221	---GASNTITTSQNVDEVDVITVPIFGESSEDDYDLQGVVGEA-----VPPSV	1269
Db	1286	GVESAQAQKQAQAOYAEEDYKVALPLFGNDDGEE--DQVTTGEAESAEPAILVPAG	1344
Qy	1270	IDNLLKSTENEEVLYLKLPLAGYRSBLKOLENNVTENVANKDLNSRFNKRENFVYL	1329
Db	1345	I-----SDYDVVYLKPLAGMYKTKIKKOLENHVAVAFNTITDMLDSLKKRNFLEV	1366
Qy	1330	ESDLPYKDKLSSVYVADKPYFLNKKRDKPLSSVNYTKSDIDPDINPANDVLGYKIL	1389
Db	1397	NSDINPKYSPSGEYIIKDPYKLLDEKKKKLLGSKITYGASIDKDLATANDGVITYYKM	1456
Qy	1390	SEKYKSDL-----DSIKKYIND-----KQGENEKLPLN	1419
Db	1457	GELYKTHLTVAVNEVKVLEADIKAEDDKIKIGSDSTKTEQTQSWAKKAELEKYLPLN	1516
Qy	1420	NIELTKYTVNDKIDLEYIHLEAKVNLVYTEKSNVEYKIKELNYLTKTIDOKLADFFKNNNF	1479
Db	1517	SLOEYSELSKAVTYTIDNLKRYINNCOLEKRAELTVKKIADYNNKMEKELEYKK---	1572
Qy	1480	VGIDSLSTDYNNHNNLLTKFLSTGVMFENLAKTVLSLNLGNLQGMNLISOHCVYKQCPQ	1539
Db	1573	---SEKNEVXKSSQLLEKLMKSKLILENBSKELISLQVLLVOTQLTMSSEHFCIDITNVPD	1628
Qy	1540	NSGCFRHLDEREBCCKLLNYKQEGDQCVENPNTCENNAGGCDADAKTEEDSGSNGKRI	1599
Db	1630	NAACYRFLDGMHEMRCLLTFKEBGGKCVGSNVTCTKDNNGGCAPEAECKMTDS---NKI	1685
Qy	1600	TCECTKPDSTPLPDGIFGCSSNFIIGISFLIIMLIL	1635
Db	1686	VCKCTKEGSEPLFEGYVCCSSSFLSFLIIMLFL	1721
RESULT	8		
Q9TYG1	Q9TYG1	PRELIMINARY: PRT: 539 AA.	
AC	Q9TYG1:		
DT	01-MAY-2000 (T-EMBLrel. 13, Created)		
DT	01-MAY-2000 (T-EMBLrel. 13, Last sequence update)		
DT	01-JUN-2000 (T-EMBLrel. 14, Last annotation update)		
DE	MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).		
GN	MSP1.		
OS	Plasmodium falciparum.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
CC	NCBI_TaxID=5833;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=93295445; PubMed=8515786;		
RA	Jongwitties S., Tanabe K., Kanbara H.;		
RT	"Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins(MSP1) of Plasmodium falciparum from		

RT field isolates.
 RL Mol. Biochem. Parasitol. 59:95-100(1993).
 DR EMBL: D13358; BAA02619.1;
 DR InterPro: IPR000561;
 DR InterPro: IPR001245;
 DR Pfam: PF00008; EGF; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW Merozoite; EGF-like domain.
 FT NON_TER
 SO SEQUENCE 539 AA; 61075 MW; C70C2E10DEC4A101 CRC64;

Query Match 33.3%; Score 2801; DB 5; Length 539;
 Best Local Similarity 100.0%; Pred. No. 2.2e-96;
 Matches 539; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1101 PLKLTSESIQEDNYASLENFKVLSKLEGLKDKNLNEKKKLSYSSGLHLLAELEKEV 1160
 DB 1 PLKLTSESIQEDNYASLENFKVLSKLEGLKDKNLNEKKKLSYSSGLHLLAELEKEV 60
 QY 1161 IKNNYTGNSPSENNVDVNNNALESYKFKFLPEGTDAVAVSSGSDTLEOSOPKKPASTHV 1220
 DB 61 IKNNYTGNSPSENNVDVNNNALESYKFKFLPEGTDAVAVSSGSDTLEOSOPKKPASTHV 120
 QY 1221 GAESNTITTSQNVDEVDVIVPIFGSEEDYDGLGVVTGEAVTPSVIDNLSKIENE 1280
 DB 121 GAESNTITTSQNVDEVDVIVPIFGSEEDYDGLGVVTGEAVTPSVIDNLSKIENE 180
 QY 1281 YEVLTKPLAGVYRSKLENNVWFENVNKKDILNSFRNREKNVLESDELIPYKDLT 1340
 DB 181 YEVLTKPLAGVYRSKLENNVWFENVNKKDILNSFRNREKNVLESDELIPYKDLT 240
 QY 1341 SSNVVVKDPYKFLNKEKDKFLSSYNYIKDSIDPIDINADVLYGKYLSEKYSDDLDSI 1400
 DB 241 SSNVVVKDPYKFLNKEKDKFLSSYNYIKDSIDPIDINADVLYGKYLSEKYSDDLDSI 300
 QY 1401 KKYINDKOGENEKYLPLNNIETLYKTVNDRIDLFVHLEAKVLNITYEKSNEVKIKEL 1460
 DB 301 KKYINDKOGENEKYLPLNNIETLYKTVNDRIDLFVHLEAKVLNITYEKSNEVKIKEL 360
 QY 1461 NYLKTIDQKLADFKKNNFVGIADLSTDYNNHNNLTKFLSGWPFENLAKTVLSMLDGN 1520
 DB 361 NYLKTIDQKLADFKKNNFVGIADLSTDYNNHNNLTKFLSGWPFENLAKTVLSMLDGN 420
 QY 1521 LOGMLNISOHCYKCKOPGSCFRHLDERECKCLNWKOGDGCVENPPTCENNNG 1580
 DB 421 LOGMLNISOHCYKCKOPGSCFRHLDERECKCLNWKOGDGCVENPPTCENNNG 480
 QY 1581 CDADAKCTEEDSGSNGKKTCECTKPDSPYPLFDGIFCSSNPLGISFLILMLILYSFI 1639
 DB 481 CDADAKCTEEDSGSNGKKTCECTKPDSPYPLFDGIFCSSNPLGISFLILMLILYSFI 539

RESULT 9
 ID 025972 PRELIMINARY; PRT: 539 AA.
 AC 025972;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, last annotation update)
 DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
 GN MSP1.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93295445; PubMed=8515786;
 RA Jongwutives S., Tanabe K., Kanbara H.;
 RT "Sequence conservation in the C-terminal part of the precursor to the
 major merozoite surface proteins (MSP1) of Plasmodium falciparum from
 field isolates."
 RL Mol. Biochem. Parasitol. 59:95-100(1993).

DR EMBL: D13360; BAA02621.1;
 DR InterPro: IPR000561;
 DR InterPro: IPR001245;
 DR Pfam: PF00008; EGF; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW Merozoite; EGF-like domain.
 FT NON_TER
 SO SEQUENCE 539 AA; 61076 MW; 72186F1412C57D1C CRC64;

Query Match 33.2%; Score 2798; DB 5; Length 539;
 Best Local Similarity 99.8%; Pred. No. 2.8e-96;
 Matches 538; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1101 PLKLTSESIQEDNYASLENFKVLSKLEGLKDKNLNEKKKLSYSSGLHLLAELEKEV 1160
 DB 1 PLKLTSESIQEDNYASLENFKVLSKLEGLKDKNLNEKKKLSYSSGLHLLAELEKEV 60
 QY 1161 IKNNYTGNSPSENNVDVNNNALESYKFKFLPEGTDAVAVSSGSDTLEOSOPKKPASTHV 1220
 DB 61 IKNNYTGNSPSENNVDVNNNALESYKFKFLPEGTDAVAVSSGSDTLEOSOPKKPASTHV 120
 QY 1221 GAESNTITTSQNVDEVDVIVPIFGSEEDYDGLGVVTGEAVTPSVIDNLSKIENE 1280
 DB 121 GAESNTITTSQNVDEVDVIVPIFGSEEDYDGLGVVTGEAVTPSVIDNLSKIENE 180
 QY 1281 YEVLTKPLAGVYRSKLENNVWFENVNKKDILNSFRNREKNVLESDELIPYKDLT 1340
 DB 181 YEVLTKPLAGVYRSKLENNVWFENVNKKDILNSFRNREKNVLESDELIPYKDLT 240
 QY 1341 SSNVVVKDPYKFLNKEKDKFLSSYNYIKDSIDPIDINADVLYGKYLSEKYSDDLDSI 1400
 DB 241 SSNVVVKDPYKFLNKEKDKFLSSYNYIKDSIDPIDINADVLYGKYLSEKYSDDLDSI 300
 QY 1401 KKYINDKOGENEKYLPLNNIETLYKTVNDRIDLFVHLEAKVLNITYEKSNEVKIKEL 1460
 DB 301 KKYINDKOGENEKYLPLNNIETLYKTVNDRIDLFVHLEAKVLNITYEKSNEVKIKEL 360
 QY 1461 NYLKTIDQKLADFKKNNFVGIADLSTDYNNHNNLTKFLSGWPFENLAKTVLSMLDGN 1520
 DB 361 NYLKTIDQKLADFKKNNFVGIADLSTDYNNHNNLTKFLSGWPFENLAKTVLSMLDGN 420
 QY 1521 LOGMLNISOHCYKCKOPGSCFRHLDERECKCLNWKOGDGCVENPPTCENNNG 1580
 DB 421 LOGMLNISOHCYKCKOPGSCFRHLDERECKCLNWKOGDGCVENPPTCENNNG 480
 QY 1581 CDADAKCTEEDSGSNGKKTCECTKPDSPYPLFDGIFCSSNPLGISFLILMLILYSFI 1639
 DB 481 CDADAKCTEEDSGSNGKKTCECTKPDSPYPLFDGIFCSSNPLGISFLILMLILYSFI 539

RESULT 10
 ID 025966 PRELIMINARY; PRT: 539 AA.
 AC 025966;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, last annotation update)
 DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
 GN MSP1.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93295445; PubMed=8515786;
 RA Jongwutives S., Tanabe K., Kanbara H.;
 RT "Sequence conservation in the C-terminal part of the precursor to the
 major merozoite surface proteins (MSP1) of Plasmodium falciparum from
 field isolates."
 RL EMBL: D13357; BAA02618.1;
 DR InterPro: IPR000561;

DR InterPro: IPR001245; -
 DR Pfam: PF00008; EGF; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW Merozoite; EGF-like domain.
 FT NON_TER
 SQ SEQUENCE 539 AA: 61144 MW: 787DE90CID0ACDE7 CRC64;

Query Match 33.2%; Score 2796; DB 5: Length 539;
 Best Local Similarity 99.8%; Pred. No. 3.3e-96;
 Matches 538; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1101 PLKLTSESIQTEDNYASLENFKVLSKLEGLKDLNLEKKKLSYLSGHLHLLAELEKEV 1160
 DB 1 PLKLTSESIQTEDNYASLENFKVLSKLEGLKDLNLEKKKLSYLSGHLHLLAELEKEV 60
 QY 1161 IKNNKNTGNSPSENNTDVNNALESYKKFLPREGTDAVAVSSGSSTLQSQPKKPASTHV 1220
 DB 61 IKNNKNTGNSPSENNTDVNNALESYKKFLPREGTDAVAVSSGSSTLQSQPKKPASTHV 120
 QY 1221 GAESNTITTSQNVDEVDVITVPIFGSESEDDYDGLGVVTCGEAVTPSVIDNLSKIENE 1280
 DB 121 GAESNTITTSQNVDEVDVITVPIFGSESEDDYDGLGVVTCGEAVTPSVIDNLSKIENE 180
 QY 1281 YEVLTKPLAGYRSLSKQLENNWTFVNVNKKDILNSFRNKRNFKNVLESDDLIPYKDLT 1340
 DB 181 YEVLTKPLAGYRSLSKQLENNWTFVNVNKKDILNSFRNKRNFKNVLESDDLIPYKDLT 240
 QY 1341 SSNNVVKDPYKFLNKEKDKFLSSYNYTKDSITDIDINADVLYGKYLSEKYSDDLDSI 1400
 DB 241 SSNNVVKDPYKFLNKEKDKFLSSYNYTKDSITDIDINADVLYGKYLSEKYSDDLDSI 300
 QY 1401 KKYINDKOGENEKYLPLFNNIETLYKTVNDRIDLFVHLEAKVLYTYEKSNVKIKEL 1460
 DB 301 KKYINDKOGENEKYLPLFNNIETLYKTVNDRIDLFVHLEAKVLYTYEKSNVKIKEL 360
 QY 1461 NYLKTIDQKLADFKNKNNFVGIADLSTIDYNNHNNLLTKFLSGWFEENLAKTVLSNLDGN 1520
 DB 361 NYLKTIDQKLADFKNKNNFVGIADLSTIDYNNHNNLLTKFLSGWFEENLAKTVLSNLDGN 420
 QY 1521 LOGMNTISOHCYKQCKQNSGCFRHLDERECKCLLYKQEGSKCVENPPTCENNNG 1580
 DB 421 LOGMNTISOHCYKQCKQNSGCFRHLDERECKCLLYKQEGSKCVENPPTCENNNG 480
 QY 1581 CDADAKCTEEDSGSNGKKTCECTKPDSPLEDFGFCSSNPLGISFLIIMLILYSFI 1639
 DB 481 CDADAKCTEEDSGSNGKKTCECTKPDSPLEDFGFCSSNPLGISFLIIMLILYSFI 539

RESULT 11

ID 025973 PRELIMINARY: PRT: 539 AA.

AC 025973;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
 GN MSP1.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93295445; PubMed=8515786;
 RA Jongwutives S., Tanabe K., Kanbara H.;
 RT "Sequence conservation in the C-terminal part of the precursor to the
 RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
 RT field isolates."
 RL Mol. Biochem. Parasitol. 59:95-100(1993).
 DR EMBL: D13361; BAA02622.1; -
 DR InterPro: IPR000561; -
 DR InterPro: IPR001245; -
 DR Pfam: PF00008; EGF; 1.

DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW Merozoite; EGF-like domain.
 FT NON_TER
 SQ SEQUENCE 539 AA: 61047 MW: 105C2E02974FB398 CRC64;

Query Match 33.2%; Score 2795; DB 5: Length 539;
 Best Local Similarity 99.8%; Pred. No. 3.6e-96;
 Matches 538; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1101 PLKLTSESIQTEDNYASLENFKVLSKLEGLKDLNLEKKKLSYLSGHLHLLAELEKEV 1160
 DB 1 PLKLTSESIQTEDNYASLENFKVLSKLEGLKDLNLEKKKLSYLSGHLHLLAELEKEV 60
 QY 1161 IKNNKNTGNSPSENNTDVNNALESYKKFLPREGTDAVAVSSGSSTLQSQPKKPASTHV 1220
 DB 61 IKNNKNTGNSPSENNTDVNNALESYKKFLPREGTDAVAVSSGSSTLQSQPKKPASTHV 120
 QY 1221 GAESNTITTSQNVDEVDVITVPIFGSESEDDYDGLGVVTCGEAVTPSVIDNLSKIENE 1280
 DB 121 GAESNTITTSQNVDEVDVITVPIFGSESEDDYDGLGVVTCGEAVTPSVIDNLSKIENE 180
 QY 1281 YEVLTKPLAGYRSLSKQLENNWTFVNVNKKDILNSFRNKRNFKNVLESDDLIPYKDLT 1340
 DB 181 YEVLTKPLAGYRSLSKQLENNWTFVNVNKKDILNSFRNKRNFKNVLESDDLIPYKDLT 240
 QY 1341 SSNNVVKDPYKFLNKEKDKFLSSYNYTKDSITDIDINADVLYGKYLSEKYSDDLDSI 1400
 DB 241 SSNNVVKDPYKFLNKEKDKFLSSYNYTKDSITDIDINADVLYGKYLSEKYSDDLDSI 300
 QY 1401 KKYINDKOGENEKYLPLFNNIETLYKTVNDRIDLFVHLEAKVLYTYEKSNVKIKEL 1460
 DB 301 KKYINDKOGENEKYLPLFNNIETLYKTVNDRIDLFVHLEAKVLYTYEKSNVKIKEL 360
 QY 1461 NYLKTIDQKLADFKNKNNFVGIADLSTIDYNNHNNLLTKFLSGWFEENLAKTVLSNLDGN 1520
 DB 361 NYLKTIDQKLADFKNKNNFVGIADLSTIDYNNHNNLLTKFLSGWFEENLAKTVLSNLDGN 420
 QY 1521 LOGMNTISOHCYKQCKQNSGCFRHLDERECKCLLYKQEGSKCVENPPTCENNNG 1580
 DB 421 LOGMNTISOHCYKQCKQNSGCFRHLDERECKCLLYKQEGSKCVENPPTCENNNG 480
 QY 1581 CDADAKCTEEDSGSNGKKTCECTKPDSPLEDFGFCSSNPLGISFLIIMLILYSFI 1639
 DB 481 CDADAKCTEEDSGSNGKKTCECTKPDSPLEDFGFCSSNPLGISFLIIMLILYSFI 539

RESULT 12

ID 025976 PRELIMINARY: PRT: 539 AA.

AC 025976;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
 GN MSP1.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93295445; PubMed=8515786;
 RA Jongwutives S., Tanabe K., Kanbara H.;
 RT "Sequence conservation in the C-terminal part of the precursor to the
 RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
 RT field isolates."
 RL Mol. Biochem. Parasitol. 59:95-100(1993).
 DR EMBL: D13362; BAA02623.1; -
 DR InterPro: IPR000561; -
 DR InterPro: IPR001245; -
 DR Pfam: PF00008; EGF; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW Merozoite; EGF-like domain.

FT NON_TER 1 1
SQ SEQUENCE 539 AA; 61045 MW; 8BF9C64322E9A778 CRC64;

Query Match 33.2%; Score 2794; DB 5; Length 539;
Best Local Similarity 99.8%; Pred. No. 3.9e-96;
Matches 538; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1101 PLKLTSEESIOTEDNYASLENFKVLSKLEGLKDKNLNEKKKLSYSSGLHLLIAELKEV 1160
DB 1 PLKLTSEESIOTEDNYASLENFKVLSKLEGLKDKNLNEKKKLSYSSGLHLLIAELKEV 60
QY 1161 IKNNYTGNSPSENNNTDVNNALLESYKKFLPGCTDVATVVSSEGSPTLEQSPKPKASTHV 1220
DB 61 IKNNYTGNSPSENNNTDVNNALLESYKKFLPGCTDVATVVSSEGSPTLEQSPKPKASTHV 120
QY 1221 GAESNTITTSQNVDEVDVIVPIFGSESEDYDGLGVVGTGEAVTPSVINDILISKIENE 1280
DB 121 GAESNTITTSQNVDEVDVIVPIFGSESEDYDGLGVVGTGEAVTPSVINDILISKIENE 180
QY 1281 YEVLTKPLAGVYSRLKKOLENNVTFNVNWKDILNSRFNRKRENFKNVLESDDLIPYKDLT 1340
DB 181 YEVLTKPLAGVYSRLKKOLENNVTFNVNWKDILNSRFNRKRENFKNVLESDDLIPYKDLT 240
QY 1341 SSNVVVKDPYKFLNKKERDKFLSSYNYIKDSIDTIDINPANDVLGYKILSEKYSDDLST 1400
DB 241 SSNVVVKDPYKFLNKKERDKFLSSYNYIKDSIDTIDINPANDVLGYKILSEKYSDDLST 300
QY 1401 KKYINDKOGENEKYLPLNNIETLYKTVNDKIDLFVHLEAKVNLVTEKSNVEVKIKEL 1460
DB 301 KKYINDKOGENEKYLPLNNIETLYKTVNDKIDLFVHLEAKVNLVTEKSNVEVKIKEL 360
QY 1461 NYLKTIDOKLADFKNNNFVGIADLSTDYNNHNNLTFLSTGMVFENLAKTVLSMLDGN 1520
DB 361 NYLKTIDOKLADFKNNNFVGIADLSTDYNNHNNLTFLSTGMVFENLAKTVLSMLDGN 420
QY 1521 LOGMLNISQHCYKQCPQNSGCFRHLDERECKCLLNYKQEGDKCVENPPTCENNNG 1580
DB 421 LOGMLNISQHCYKQCPQNSGCFRHLDERECKCLLNYKQEGDKCVENPPTCENNNG 480
QY 1581 CDADAKCTEEDSGSGNGKKTCECTKPDSPYPLFDGIFGSSSNFLGISFLIIMLILYSPI 1639
DB 481 CDADAKCTEEDSGSGNGKKTCECTKPDSPYPLFDGIFGSSSNFLGISFLIIMLILYSPI 539

RESULT 13
Q25981
ID Q25981 PRELIMINARY; PRT; 539 AA.
AC Q25981;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwitties S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13363; BAA02624.1; -
DR InterPro: IPR000561; -
DR InterPro: IPR001245; -
DR Pfam: PF00008; EGF.1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1 1
SQ SEQUENCE 539 AA; 61046 MW; 3EBDB7473EB87B65 CRC64;

Query Match 33.1%; Score 2791; DB 5; Length 539;
Best Local Similarity 99.6%; Pred. No. 5.1e-96;
Matches 537; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1101 PLKLTSEESIOTEDNYASLENFKVLSKLEGLKDKNLNEKKKLSYSSGLHLLIAELKEV 1160
DB 1 PLKLTSEESIOTEDNYASLENFKVLSKLEGLKDKNLNEKKKLSYSSGLHLLIAELKEV 60
QY 1161 IKNNYTGNSPSENNNTDVNNALLESYKKFLPGCTDVATVVSSEGSPTLEQSPKPKASTHV 1220
DB 61 IKNNYTGNSPSENNNTDVNNALLESYKKFLPGCTDVATVVSSEGSPTLEQSPKPKASTHV 120
QY 1221 GAESNTITTSQNVDEVDVIVPIFGSESEDYDGLGVVGTGEAVTPSVINDILISKIENE 1280
DB 121 GAESNTITTSQNVDEVDVIVPIFGSESEDYDGLGVVGTGEAVTPSVINDILISKIENE 180
QY 1281 YEVLTKPLAGVYSRLKKOLENNVTFNVNWKDILNSRFNRKRENFKNVLESDDLIPYKDLT 1340
DB 181 YEVLTKPLAGVYSRLKKOLENNVTFNVNWKDILNSRFNRKRENFKNVLESDDLIPYKDLT 240
QY 1341 SSNVVVKDPYKFLNKKERDKFLSSYNYIKDSIDTIDINPANDVLGYKILSEKYSDDLST 1400
DB 241 SSNVVVKDPYKFLNKKERDKFLSSYNYIKDSIDTIDINPANDVLGYKILSEKYSDDLST 300
QY 1401 KKYINDKOGENEKYLPLNNIETLYKTVNDKIDLFVHLEAKVNLVTEKSNVEVKIKEL 1460
DB 301 KKYINDKOGENEKYLPLNNIETLYKTVNDKIDLFVHLEAKVNLVTEKSNVEVKIKEL 360
QY 1461 NYLKTIDOKLADFKNNNFVGIADLSTDYNNHNNLTFLSTGMVFENLAKTVLSMLDGN 1520
DB 361 NYLKTIDOKLADFKNNNFVGIADLSTDYNNHNNLTFLSTGMVFENLAKTVLSMLDGN 420
QY 1521 LOGMLNISQHCYKQCPQNSGCFRHLDERECKCLLNYKQEGDKCVENPPTCENNNG 1580
DB 421 LOGMLNISQHCYKQCPQNSGCFRHLDERECKCLLNYKQEGDKCVENPPTCENNNG 480
QY 1581 CDADAKCTEEDSGSGNGKKTCECTKPDSPYPLFDGIFGSSSNFLGISFLIIMLILYSPI 1639
DB 481 CDADAKCTEEDSGSGNGKKTCECTKPDSPYPLFDGIFGSSSNFLGISFLIIMLILYSPI 539

RESULT 14
Q25984
ID Q25984 PRELIMINARY; PRT; 539 AA.
AC Q25984;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwitties S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13356; BAA02617.1; -
DR InterPro: IPR000561; -
DR InterPro: IPR001245; -
DR Pfam: PF00008; EGF.1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1 1
SQ SEQUENCE 539 AA; 61114 MW; 3788015F3127CB9E CRC64;

Query Match 33.1%; Score 2789; DB 5; Length 539;
 Best Local Similarity 99.6%; Pred. No. 6e-96;
 Matches 537; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1101 PLKLTSEESIOTEDNYASLENFVKLSKLEGLKDNLNIEKKKLSYSSGLHHLIAELKEV 1160
 |||||||
 DB 1 PLKLTSEESIOTEDNYASLENFVKLSKLEGLKDNLNIEKKKLSYSSGLHHLIAELKEV 60
 QY 1161 IKNNNYTGNPSSENNNTDVNNALLESYKKFLPEGTDAVATVSSGSDTLEQSPKKPASTHV 1220
 |||||||
 DB 61 IKNNNYTGNPSVNNNTDVNNALLESYKKFLPEGTDAVATVSSGSDTLEQSPKKPASTHV 120
 QY 1221 GAESNTTTSQNVDEVDVITVPIFGSEEDYDLDGQVYTGAVTSPVINDILSKIENE 1280
 |||||||
 DB 121 GAESNTTTSQNVDEVDVITVPIFGSEEDYDLDGQVYTGAVTSPVINDILSKIENE 180
 QY 1281 YEVLTKPLAGVYRSLSKKOLENNVMTFNVKDILNSRFNRKREPKNVLESDLIPYKDLT 1340
 |||||||
 DB 181 YEVLTKPLAGVYRSLSKKOLENNVMTFNVKDILNSRFNRKREPKNVLESDLIPYKDLT 240
 QY 1341 SSNVVVDPYKFLNKEKRDKLSSYNYIKDSIDPDINFANDVILGYKILSEKYSDDLSDSI 1400
 |||||||
 DB 241 SSNVVVDPYKFLNKEKRDKLSSYNYIKDSIDPDINFANDVILGYKILSEKYSDDLSDSI 300
 QY 1401 KKYINDKOGENEKYLPELNNIETLYKTVNDKIDLFVHLEAKVINTYEEKSNVEVKIKEL 1460
 |||||||
 DB 301 KKYINDKOGENEKYLPELNNIETLYKTVNDKIDLFVHLEAKVINTYEEKSNVEVKIKEL 360
 QY 1461 NYLKTIOOKLADFKKNNNFVGIADLSTDYNNHNLTKFLSTGMPFENLAKTVLSNLDGN 1520
 |||||||
 DB 361 NYLKTIOOKLADFKKNNNFVGIADLSTDYNNHNLTKFLSTGMPFENLAKTVLSNLDGN 420
 QY 1521 LOGMLNISOHCQVKKOCPONSGCFRHLDERECKCLNLYKQEGDKCVENPPTCENNNGG 1580
 |||||||
 DB 421 LOGMLNISOHCQVKKOCPONSGCFRHLDERECKCLNLYKQEGDKCVENPPTCENNNGG 480
 QY 1581 CDADAKCTEEDSGSNGKKITCECTKPDSPYPLFDGIFGSSNPLGISFLILMLILYSFI 1639
 |||||||
 DB 481 CDADAKCTEEDSGSNGKKITCECTKPDSPYPLFDGIFGSSNPLGISFLILMLILYSFI 539

RESULT 15
 Q25971 PRELIMINARY; PRT: 539 AA.

ID Q25971
 AC Q25971;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
 DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
 GN MSP1.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9329445; PubMed=8515786;
 RA Jongwutives S., Tanabe K., Kanbara H.;
 RT "Sequence conservation in the C-terminal part of the precursor to the
 RT major merozoite surface proteins (MSP) of Plasmodium falciparum from
 RT field isolates";
 RL Mol. Biochem. Parasitol. 59:95-100(1993).
 DR EMBL; D13359; BAA02620.1; -;
 DR InterPro: IPR000561; -;
 DR InterPro: IPR001245; -;
 DR Pfam: PF00008; EGF.1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR. 1.
 KW Merozoite; EGF-like domain.
 FT NON_TER 1
 SQ SEQUENCE 539 AA; 61017 MW; 5CA9C651BB62B5E1 CRC64;

Matches 537; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1101 PLKLTSEESIOTEDNYASLENFVKLSKLEGLKDNLNIEKKKLSYSSGLHHLIAELKEV 1160
 |||||||
 DB 1 PLKLTSEESIOTEDNYASLENFVKLSKLEGLKDNLNIEKKKLSYSSGLHHLIAELKEV 60
 QY 1161 IKNNNYTGNPSSENNNTDVNNALLESYKKFLPEGTDAVATVSSGSDTLEQSPKKPASTHV 1220
 |||||||
 DB 61 IKNNNYTGNPSVNNNTDVNNALLESYKKFLPEGTDAVATVSSGSDTLEQSPKKPASTHV 120
 QY 1221 GAESNTTTSQNVDEVDVITVPIFGSEEDYDLDGQVYTGAVTSPVINDILSKIENE 1280
 |||||||
 DB 121 GAESNTTTSQNVDEVDVITVPIFGSEEDYDLDGQVYTGAVTSPVINDILSKIENE 180
 QY 1281 YEVLTKPLAGVYRSLSKKOLENNVMTFNVKDILNSRFNRKREPKNVLESDLIPYKDLT 1340
 |||||||
 DB 181 YEVLTKPLAGVYRSLSKKOLENNVMTFNVKDILNSRFNRKREPKNVLESDLIPYKDLT 240
 QY 1341 SSNVVVDPYKFLNKEKRDKLSSYNYIKDSIDPDINFANDVILGYKILSEKYSDDLSDSI 1400
 |||||||
 DB 241 SSNVVVDPYKFLNKEKRDKLSSYNYIKDSIDPDINFANDVILGYKILSEKYSDDLSDSI 300
 QY 1401 KKYINDKOGENEKYLPELNNIETLYKTVNDKIDLFVHLEAKVINTYEEKSNVEVKIKEL 1460
 |||||||
 DB 301 KKYINDKOGENEKYLPELNNIETLYKTVNDKIDLFVHLEAKVINTYEEKSNVEVKIKEL 360
 QY 1461 NYLKTIOOKLADFKKNNNFVGIADLSTDYNNHNLTKFLSTGMPFENLAKTVLSNLDGN 1520
 |||||||
 DB 361 NYLKTIOOKLADFKKNNNFVGIADLSTDYNNHNLTKFLSTGMPFENLAKTVLSNLDGN 420
 QY 1521 LOGMLNISOHCQVKKOCPONSGCFRHLDERECKCLNLYKQEGDKCVENPPTCENNNGG 1580
 |||||||
 DB 421 LOGMLNISOHCQVKKOCPONSGCFRHLDERECKCLNLYKQEGDKCVENPPTCENNNGG 480
 QY 1581 CDADAKCTEEDSGSNGKKITCECTKPDSPYPLFDGIFGSSNPLGISFLILMLILYSFI 1639
 |||||||
 DB 481 CDADAKCTEEDSGSNGKKITCECTKPDSPYPLFDGIFGSSNPLGISFLILMLILYSFI 539

Search completed: October 27, 2001, 15:55:53
 Job time: 273 sec

Query Match 33.1%; Score 2788; DB 5; Length 539;
 Best Local Similarity 99.6%; Pred. No. 6.5e-96;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2001, 15:53:10 ; Search time 23.62 Seconds

(without alignments)
2377.000 Million cell updates/sec

Title: US-09-269-874-3

Perfect score: 8424

Sequence: 1 MKRIIFLCSEFLFIINTOCV.....SNFLGISPLLLMLILYSFI 1639

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8424	100.0	1639	1 MSP1_PLAFW	P04933 plasmodium
2	8144.5	96.7	1630	1 MSP1_PLAFK	P04932 plasmodium
3	5136.5	61.0	1726	1 MSP1_PLAFB	P04934 plasmodium
4	5127.5	60.9	1726	1 MSP1_PLAFB	P50495 plasmodium
5	5069	60.2	1701	1 MSP1_PLAFM	P08569 plasmodium
6	5065	60.1	1701	1 MSP1_PLAFB	P13819 plasmodium
7	4854.5	57.6	1682	1 MSP1_PLAFB	P19598 plasmodium
8	2548.5	30.3	1772	1 MSP1_PLAFD	P13828 plasmodium
9	1187	14.1	233	1 MSP1_PLAFD	P13827 plasmodium
10	1096.5	13.0	281	1 MSP1_PLAFN	P13820 plasmodium
11	501.5	6.0	1957	1 YD86_SCHPO	Q10411 schistosom
12	486	5.8	1790	1 USO1_YEAST	P23386 saccharomyc
13	477	5.7	1869	1 RBP1_PLAFB	Q00798 plasmodium
14	474.5	5.6	1875	1 MLPI_YEAST	Q02455 saccharomyc
15	457.5	5.4	1251	1 RBP2_PLAFB	Q00799 plasmodium
16	443.5	5.3	2663	1 CENE_HUMAN	Q02224 homo sapien
17	440	5.2	1805	1 HMM2_MYGE	P47460 mycoplasma
18	424	5.0	2230	1 GOG4_HUMAN	Q13439 homo sapien
19	423	5.0	1679	1 Y109_YEAST	P40457 saccharomyc
20	421.5	5.0	1928	1 MYS1_YEAST	P08964 saccharomyc
21	405	4.8	2022	1 ANTL1_ONCVO	P13249 onchocerca
22	404	4.8	2116	1 MYS2_DICDI	P08799 dictyosteli
23	393	4.7	1312	1 RA50_YEAST	P12753 saccharomyc
24	386.5	4.6	2748	1 NUM1_YEAST	Q00402 saccharomyc
25	378	4.5	1818	1 HMM2_MYCPN	P75471 mycoplasma
26	361	4.3	3660	1 DMD_CHICK	P15533 gallus galli
27	358.5	4.3	2349	1 TPR_HUMAN	P12270 homo sapien
28	355.5	4.2	3210	1 CENE_HUMAN	P49454 homo sapien
29	351.5	4.2	1935	1 MYSB_RAT	P02564 rattus norv
30	349.5	4.1	1935	1 MYSB_HUMAN	P12883 homo sapien
31	349.5	4.1	1935	1 YLJ2_CAEEL	P14367 caenorhabdi
32	349	4.1	1939	1 MYSB_MESAU	P13539 mesocricetu
33	348	4.1	1940	1 MYSB_CHICK	P02565 gallus galli

34	347	4.1	1225	1	SMC1_YEAST	P32908 saccharomyc
35	345.5	4.1	2649	1	BPAL_HUMAN	O03001 homo sapien
36	344.5	4.1	1934	1	MYSB_MESAU	P13540 mesocricetu
37	344	4.1	2136	1	YCF2_MARPO	P09975 marichantia
38	343	4.1	1939	1	MYSB_HUMAN	P13533 homo sapien
39	341.5	4.1	1938	1	MYSB_MOUSE	Q02566 mus musculu
40	341.5	4.1	1978	1	MYSB_CHICK	P10587 gallus galli
41	338	4.0	1169	1	EX5B_BORBU	O51578 borrelia bu
42	338	4.0	1938	1	MYSB_RAT	P02563 rattus norv
43	337.5	4.0	1539	1	Y373_HUMAN	O15078 homo sapien
44	336.5	4.0	1427	1	REST_HUMAN	P30622 homo sapien
45	335	4.0	3685	1	DMD_HUMAN	P11532 homo sapien

ALIGNMENTS

RESULT	ID	MSPI_PLAFW	STANDARD:	PRT:	1639 AA.
AC	P04933:	13-AUG-1987 (Rel. 05, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	01-FEB-1996 (Rel. 33, Last annotation update)				
DE	MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)				
DE	(PMMSA) (P195).				
GN	MSP-1.				
OS	Plasmodium falciparum (isolate Wellcome).				
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
OX	NCBI_TaxID=5848;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=86014355; PubMed=2995820;				
RA	Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,				
RA	Nicholls S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,				
RA	Freeman R.R.;				
RT	"Primary structure of the precursor to the three major surface				
RT	antigens of Plasmodium falciparum merozoites.",				
RN	Nature 317:270-273(1985).				
RL	[2]				
RP	REVISIONS.				
RA	Holder A.A.;				
RL	Submitted (MAR-1991) to the EMBL/Genbank/DBJ databases.				
CC	- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR				
CC	(POTENTIAL).				
CC	- PPM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42				
CC	KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF				
CC	MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; X02919; CAA26676.1; -				
DR	PIR; A24594; A24594.				
DR	InterPro: IPR000561; -				
DR	Pfam: PF00008; BGF; 1.				
KW	Maternal; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;				
KW	Transmembrane; GPI-anchor.				
FT	SIGNAL	1	19		POTENTIAL.
FT	CHAIN	20	1639		MEROZOITE SURFACE PROTEIN 1.
FT	CARBOHYD	116	116		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	268	268		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	764	764		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	768	768		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	783	783		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	844	844		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	920	920		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	964	964		N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1058 1058 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1445 1445 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1526 1526 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 1639 AA: 187618 MM: 2C235B616C87F6E CRC64:

Query Match 100.0%; Score 8424; DB 1; Length 1639;
 Best Local Similarity 100.0%; Pred. No. 3.7e-255;
 Matches 1639; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKIIFELCSLEFIINTQCVTHESIYQELVKKLEALEDAVLTFGYSLFQKEKKVNLNGTSGT 60
 DB 1 MKIIFELCSLEFIINTQCVTHESIYQELVKKLEALEDAVLTFGYSLFQKEKKVNLNGTSGT 60
 QY 61 AVTTSTPGSKGSVASGGSGSVASGGSVASGGSVASGGSNGNRPNPSNDS 120
 DB 61 AVTTSTPGSKGSVASGGSGSVASGGSVASGGSVASGGSNGNRPNPSNDS 120
 QY 121 DAKSYADLKHVRNYLTIKELKYPQLFDLTNHLTLCDNTHGFYLLDGYEINELLK 180
 DB 121 DAKSYADLKHVRNYLTIKELKYPQLFDLTNHLTLCDNTHGFYLLDGYEINELLK 180
 QY 181 LNFYFDLLRAKLVNDVCAVDYCOIPENLKRANELDVLKLVGYRKLDPDNIKDNYKMD 240
 DB 181 LNFYFDLLRAKLVNDVCAVDYCOIPENLKRANELDVLKLVGYRKLDPDNIKDNYKMD 240
 QY 241 YIKKNNKTIENINELIESKTIKNNKATKEEKKKIYQAOYDSTYNNKOLEPHNLIS 300
 DB 241 YIKKNNKTIENINELIESKTIKNNKATKEEKKKIYQAOYDSTYNNKOLEPHNLIS 300
 QY 301 VLEKRIDLKNNENIKELDKINELKPNPPANGSNTPTLLDKNNKIEHEKEIKI 360
 DB 301 VLEKRIDLKNNENIKELDKINELKPNPPANGSNTPTLLDKNNKIEHEKEIKI 360
 QY 361 IKFNIDSLFTDPLELEYLRKNNKIDISAKVETKESTPEPNYGVYPLSYNDINNAL 420
 DB 361 IKFNIDSLFTDPLELEYLRKNNKIDISAKVETKESTPEPNYGVYPLSYNDINNAL 420
 QY 421 NELSFQGLIMPFTYKPSKNITYDNERRKFINELIKERIKESDKKSYDRSKS 480
 DB 421 NELSFQGLIMPFTYKPSKNITYDNERRKFINELIKERIKESDKKSYDRSKS 480
 QY 481 LNDITKEEKLNLNEYDSEKFNNDIDLTNEFKMMGRSYKYVEKLHNHTFASYENK 540
 DB 481 LNDITKEEKLNLNEYDSEKFNNDIDLTNEFKMMGRSYKYVEKLHNHTFASYENK 540
 QY 541 EKLTKALKMEDYSLRNIVEREKLYKNLISKIENIEJTVENIKKDEQLFEKKITKD 600
 DB 541 EKLTKALKMEDYSLRNIVEREKLYKNLISKIENIEJTVENIKKDEQLFEKKITKD 600
 QY 601 ENKDEKILVESDIYKOVQVLLMKNIDELAKTOLLKNNELKNITVPSYKOEKOE 660
 DB 601 ENKDEKILVESDIYKOVQVLLMKNIDELAKTOLLKNNELKNITVPSYKOEKOE 660
 QY 661 PYYLILVLEKEIDKLKVPMPKVESLINEKKNITKTEGOSDNPSPEGEITGQATKPPGO 720
 DB 661 PYYLILVLEKEIDKLKVPMPKVESLINEKKNITKTEGOSDNPSPEGEITGQATKPPGO 720
 QY 721 AGSALGDSVQAQOEOQAPVPVPEAKAQPPTPAPVNNKTENVSKIDYLEKLYE 780
 DB 721 AGSALGDSVQAQOEOQAPVPVPEAKAQPPTPAPVNNKTENVSKIDYLEKLYE 780
 QY 781 FLNLSYICHKIILVSHSTNNEKILKQYKITEEESKLSCEPLDLFNINNIPIVMSMF 840
 DB 781 FLNLSYICHKIILVSHSTNNEKILKQYKITEEESKLSCEPLDLFNINNIPIVMSMF 840
 QY 841 DSLNNLSQLMEIYERKEMVNLKKNKNDKIKNLEAKKAVSTSVKTLSSSMQPLST 900
 DB 841 DSLNNLSQLMEIYERKEMVNLKKNKNDKIKNLEAKKAVSTSVKTLSSSMQPLST 900
 QY 901 PODKPEVSANDTSHSTNLSLKFENILSLGKNKNIYQELIGKSSSENYEYKILKDS 960

DB 901 PODKPEVSANDTSHSTNLSLKFENILSLGKNKNIYQELIGKSSSENYEYKILKDS 960
 QY 961 TFYNSEFNPFKSRADDTNSINDESKRKLLEEDINKKLTOLSDPLDKNNKILKLERFD 1020
 DB 961 TFYNSEFNPFKSRADDTNSINDESKRKLLEEDINKKLTOLSDPLDKNNKILKLERFD 1020
 QY 1021 KKKTVGKTKMOIKKLTLLKEOLESKLNSLNPPKHVLONEVFPNKKKAEIETENTLEN 1080
 DB 1021 KKKTVGKTKMOIKKLTLLKEOLESKLNSLNPPKHVLONEVFPNKKKAEIETENTLEN 1080
 QY 1081 TKILKHKYGLVKKYNGSSPLKLTSESIOTEDNYASLENPKVLSKLEGLKDNLEK 1140
 DB 1081 TKILKHKYGLVKKYNGSSPLKLTSESIOTEDNYASLENPKVLSKLEGLKDNLEK 1140
 QY 1141 KKLSTLSSGLHLLIAELKEVKNKNYNTGNSPENNVDVNNLSEYKKPLPGTDVATVVS 1200
 DB 1141 KKLSTLSSGLHLLIAELKEVKNKNYNTGNSPENNVDVNNLSEYKKPLPGTDVATVVS 1200
 QY 1201 EGSPTLEQSOPKKPASTHVAESNTITTSQNVDEVDVITVPFGESEEDYDLGQV 1260
 DB 1201 EGSPTLEQSOPKKPASTHVAESNTITTSQNVDEVDVITVPFGESEEDYDLGQV 1260
 QY 1261 TGEAVTPSVIDNIIISKIENEYEVLYKPLAGVYSLKKOLENNWTFENVNKKDILNSRFN 1320
 DB 1261 TGEAVTPSVIDNIIISKIENEYEVLYKPLAGVYSLKKOLENNWTFENVNKKDILNSRFN 1320
 QY 1321 KRENKKNVLESGLIPYKDLTSSNVVVKDPYKFLNKKERDKFLSSVNYTKDSIDDDINAN 1380
 DB 1321 KRENKKNVLESGLIPYKDLTSSNVVVKDPYKFLNKKERDKFLSSVNYTKDSIDDDINAN 1380
 QY 1381 DVLGYKILSEKYSDDLDSIKKYYINDKOGENEKYLPLNNIETLYKYVNDKIDLFVHLE 1440
 DB 1381 DVLGYKILSEKYSDDLDSIKKYYINDKOGENEKYLPLNNIETLYKYVNDKIDLFVHLE 1440
 QY 1441 AKVILNYYEKSNEVKIKELNLYKTIDKLDKFKNNPFGIADLSTDYNNHNNLLTKFLS 1500
 DB 1441 AKVILNYYEKSNEVKIKELNLYKTIDKLDKFKNNPFGIADLSTDYNNHNNLLTKFLS 1500
 QY 1501 TGMVFENLAKTVLSNLDLNGNLOGMNLSOHOCVKQCOONSQCFPHLDERECKOLNLYK 1560
 DB 1501 TGMVFENLAKTVLSNLDLNGNLOGMNLSOHOCVKQCOONSQCFPHLDERECKOLNLYK 1560
 QY 1561 QEGDKCVENPPTCENNNGGDADAKCTEEDSGSNGKKTCECTKPDSPPLFDGIFCSSS 1620
 DB 1561 QEGDKCVENPPTCENNNGGDADAKCTEEDSGSNGKKTCECTKPDSPPLFDGIFCSSS 1620
 QY 1621 NPLGISFLIILMLIYSFI 1639
 DB 1621 NPLGISFLIILMLIYSFI 1639

RESULT 2
 MSP1_PLAFK STANDARD; PRT: 1630 AA.
 ID MSP1_PLAFK P04932;
 AC P04932;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE MEMOZOITE SURFACE PROTEIN 1 PRECURSOR (MEMOZOITE SURFACE ANTIGENS)
 DE (PMMSA) (P190).
 GN MSP-1.
 OS Plasmodium falciparum (isolate KI / Thailand).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_Taxid=3839;
 RX MEDLINE=86136024; Pubmed=3004972;
 RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
 RA Stuenkelberg H., Bujard H.;
 RT "Polymorphism of the precursor for the major surface antigens of
 Plasmodium falciparum merozoites: studies at the genetic level.";
 RL EMBO J. 4:3823-3829(1985).

[2]
RN REVISIONS, SEQUENCE FROM N.A.
RA Pan W., Tolle R., Bujard H.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -1- PPM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03371; CAA27070.1;
DR PIR; A25120; SAZOK1.
DR InterPro: IPR000561;
DR Pfam: PF00008; EGF; 1.
KW Malaria; Merozoite; Polypotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1630 MEROZOITE SURFACE PROTEIN 1.
FT DOMAIN 67 84 TRIPEPTIDE SC(TP) REPEAT.
FT TRANSMEM 1614 1630 MEMBRANE ANCHOR.
FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 755 755 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 759 759 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 774 774 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 835 835 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 911 911 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 955 955 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1049 1049 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1156 1156 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1436 1436 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1517 1517 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1630 AA; 187289 MW; ADBDEC3CE0A46322 CRC64;
Query Match 96.7%; Score 8144.5; DR 1; Length 1630;
Best Local Similarity 97.1%; Pred. No. 1.8e-246;
Matches 1591; Conservative 11; Mismatches 28; Indels 9; Gaps 2;

Db 352 IKFNIDSLFTDPLELEYLRKKNKIDISAKVETKESTPEPEYNGVYPLSYNDINNAL 411
Qy 421 NELNSFGDLINFDYTKPEKSKNIYTDNEKKKFINIKIKIEKKIESDKKYSDBRSK 480
Db 412 NELNSFGDLINFDYTKPEKSKNIYTDNEKKKFINIKIKIEKKIESDKKYSDBRSK 471
Qy 481 LINDIKKEYEKLINLEYDSKFNNDIDLTNEKMMGKRYKYVEKILHNTFPASYENSKNL 540
Db 472 LNDITKEYEKLINLEYDSKFNNDIDLTNEKMMGKRYKYVEKILHNTFPASYENSKNL 531
Qy 541 EKLTKALYMEDYSLRNIVVEKEKILKYRNILSKITENIETIYENIKKDEQLFEKKIKD 600
Db 532 EKLTKALYMEDYSLRNIVVEKEKILKYRNILSKITENIETIYENIKKDEQLFEKKIKD 591
Qy 601 ENKPEDEKILEVSDIVKYOVQVLLMANKIDELKKTQILKNVELKHNHVPSYKOEKOE 660
Db 592 ENKPEDEKILEVSDIVKYOVQVLLMANKIDELKKTQILKNVELKHNHVPSYKOEKOE 651
Qy 661 PYLLIVLKKELIDKLKVFMPKVESLINEEKNKITEGQSDNSEPTEGBITQOATTKPQOQ 720
Db 652 PYLLIVLKKELIDKLKVFMPKVESLINEEKNKITEGQSDNSEPTEGBITQOATTKPQOQ 711
Qy 721 AGSALEGDVVOAQAOEQKQAPVPVPEPAKQVPTPPAPVNNKTEVNSKIDYLEKLYE 780
Db 712 AGSALEGDVVOAQAOEQKQAPVPVPEPAKQVPTPPAPVNNKTEVNSKIDYLEKLYE 771
Qy 781 FLNTSYICHKYLIVSHSTMEKILKYKITEEESKSSCDPLDLEFNINNNIPVMSMF 840
Db 772 FLNTSYICHKYLIVSHSTMEKILKYKITEEESKSSCDPLDLEFNINNNIPVMSMF 831
Qy 841 DSLNNSISQLEMEIYERKEMVGNLYKLNKDNKIKNLLEAKKAVSYSVKTLSSSSWQPLSLT 900
Db 832 DSLNNSISQLEMEIYERKEMVGNLYKLNKDNKIKNLLEAKKAVSYSVKTLSSSSWQPLSLT 891
Qy 901 PODKREVSANDTSHSTMNNSLSKLFENILSLGKKNNTYQOELIGKSSSENYEKLTKSD 960
Db 892 PODKREVSANDTSHSTMNNSLSKLFENILSLGKKNNTYQOELIGKSSSENYEKLTKSD 951
Qy 961 TPYNESFTNPKSKADINSINDESKRRKLEEDINKLKTQLSFDLYNKKYKLERLFD 1020
Db 952 TPYNESFTNPKSKADINSINDESKRRKLEEDINKLKTQLSFDLYNKKYKLERLFD 1011
Qy 1021 KKKTVGKTKMOKIKTLTKLEQLESKLNLSNPKHYLQNFVSFNNKKRAELAEFTENTLEN 1080
Db 1012 KKKTVGKTKMOKIKTLTKLEQLESKLNLSNPKHYLQNFVSFNNKKRAELAEFTENTLEN 1071
Qy 1081 TKILLKHYKGLVKKYNGESSPLKTSEESIOTEDNYASLENEKVLSEGLKIDNLNLEK 1140
Db 1072 TKILLKHYKGLVKKYNGESSPLKTSEESIOTEDNYASLENEKVLSEGLKIDNLNLEK 1131
Qy 1141 KKLSTYSSGHLHLIAELKEVIRKNNKYTGNSPENNVDVNNALLESYKFLPGCTQVATVVS 1200
Db 1132 KKLSTYSSGHLHLIAELKEVIRKNNKYTGNSPENNVDVNNALLESYKFLPGCTQVATVVS 1191
Qy 1201 ESSGPTLEQSQPKKRPASHVGAESNTITTSQNVDEVDVYIIVPIFGSEEDYDGLGVY 1260
Db 1192 ESSGPTLEQSQPKKRPASHVGAESNTITTSQNVDEVDVYIIVPIFGSEEDYDGLGVY 1251
Qy 1261 TGEAVTPSVINILISKITENIEVEVLTKPLAGVYSLSKQLENNNMTFVANNKIDJINSFEN 1320
Db 1252 TGEAVTPSVINILISKITENIEVEVLTKPLAGVYSLSKQLENNNMTFVANNKIDJINSFEN 1311
Qy 1321 KRENFKNVLESDDLIPYKDLTSSNNVVKDPYFLNKKERDKFLSSYNTKDSIDFDINAN 1380
Db 1312 KRENFKNVLESDDLIPYKDLTSSNNVVKDPYFLNKKERDKFLSSYNTKDSIDFDINAN 1371
Qy 1381 DVLGYKTLSEKYSKSDLSIKKYYINDKOGENKYLPLNNIETLYKYVNDKIDLEFVHLE 1440
Db 1372 DVLGYKTLSEKYSKSDLSIKKYYINDKOGENKYLPLNNIETLYKYVNDKIDLEFVHLE 1431
Qy 1441 AKVINYTEKSNVEKIKELVYKTIODKLADFKKNNNFVGIADISTDYNNNNLTFLS 1500

QY	332	NSGTPMTLDDKNKKIEHEHEKEIKETAKTITKFNIDSLFDPDLELEYLREKKNIDISAK	391
Db	359	NSGTPMTLDDKNKKIEHEHEKEIKETAKTITKFNIDSLFDPDLELEYLREKKNIDISAK	418
QY	392	VE--TKRSTEPN--EYPGVVYPLSYNDINNAL---NELMSFGDLINFPDYTPESKNIYT	445
Db	419	SQDPTKSVQJPKPVYPRNGIYVPLPLDINHSLAODKNSYDGLMNP--DTRKKNIEKIT	477
QY	446	DN--ERKFFINDEIKELIKEKKI---ESDKSYEDRSKSLNDITKEKELNIYEYSKF	500
Db	478	DNKERKTEFINNKKQJIDLEKKINTHTKEONKKLEDEYKES---KKDYELLEKEFYEMKF	533
QY	501	NNNIDLNFEMKMGKRSYXVEKTLHHNTPASYSNKHLEKLTATLAKMDEYSLRNIVV	560
Db	534	NNNDDKDVVKITISARTAYVVEKORNNKFRSSNNNSYVAVQKIKALSLYEDJSLRGKIS	593
QY	561	EKELEYKKNLSIKIENIEFLNENIKKDEOLEERK---ITDEKMPDERKILEVEDIYKV	617
Db	594	EKDPNNHYTLKTGEADIKKLTPEIKESSEKKILEKFKGLTSHSANA---LEVYDIYKL	649
QY	618	QOVQFVLLMKNIDELKKTQJLILKNVELEKHNHIVPNSYKOEKNQEPYLLVYLKKEIDKAVE	677
Db	650	QOVQVLLTIKIEDLRJTEFLNAQJLKDSTIHVBNYIKPONKREPYLLVYLKREVDKLEF	709
QY	678	MPKYESLINEEKKNIKIEGSDNSSESTBEITIGQATTTPRGQAQALBGSDVQAQAOBQ	737
Db	710	TPYKMDIKKREQAVLSS-----ITQPLVASFETTEDGGSHTHLTSQSGETEVEET	760
QY	738	KQAPRPV-----PVPVEAKAQQVTPPAPVNNKTEVANSKIDYEKLEPFLNYSYCH	789
Db	761	ETEBEYVGHITTYITILPREVYKVENSIEMKSNDSQALTKTYILKIDELTSLSYCH	820
QY	790	KYILVSHSTINEKILKOYKITKEESKLSCDPLDLFNIONNIPVMSFMDPLNNSLSQ	849
Db	821	KYILVSNSSMQOKLEYYNLTPEEENELKSCDPLDLFNIONNIPAMSYLYSDMNNDOH	880
QY	850	LFMBITYEKEVANCILYKJDKNDKIKNLEAKKY-----SISYKTLSSS-----	893
Db	881	LFEPLOYEKMIYLLHYLKRJENHAIKLEBOKQJTGTSSTSPBNTVYNAQATSHNSON	940
QY	894	--MOPLSUTPODKREVSANB---DTSHS---TNLNSLSLKFENILSLGKNINITYQEL-I	943
Db	941	QOSNASSTNONGVAVSSGAVVEESHDPLVJLSTINDLKGIVSLNLGNKTKVYNPPLTI	1000
QY	944	GQKSSSENFYKILKIDSDTFYNESFTNFVSKRADDLSLNDSEKRRKLEIKLKTJOL	1003
Db	1001	STTEMEKFEYENILKNNDTYFNDDIKQFVXSNSKVITGLF-ETQKNALINDEIKKLDJOL	1059
QY	1004	SFDLYNKYKILKLELPKPKKTTVGKTKMQIKKLTLKEQJESKLNSJNPNKHYLQNFVSFF	1066
Db	1060	SFDLYNKYKILKLELPKPKKTTVGKTKMQIKKLTLKEQJESKLNSJNPNHVLQNFVSFF	1119
QY	1064	NKKKEALIAETNLTENTKLLKHVGLVYVYVYGGESSPKLTSEESIQEDMYASLENFK	1123
Db	1120	NKKKEALIAETNLTENTKLLKHVGLVYVYVYVYGGESSPKLTSEESIQEDMYANDEKFR	1179
QY	1124	VLSKLESLKLDNINLEKKKLSYLSGGLHLHLIAELKLVYKKNVYTGNSPENNTPYNNAL	1183
Db	1180	VLSKIDTKLDNIDLHGLKRRKLSFSSGHLQITELKEVYKKNVYTGNSPENNNKRVNEMAL	1239
QY	1184	SYKFEPLPEGDIATVASE-----SGSOTLEQSQPKKPAISHVGAESNTITTT	1229
Db	1240	SYENFLEP--AKVYTVVYVPPORDYTPSPSLVARSVSGSGSRKEFTQJPTPSGLTLELQOOVO	1298
QY	1230	SQNVDDDEVADVYIYPIJGSESEEDYDQJGOVVGTEAVTPYVIOINLSIKINEVEVYKPL	1289
Db	1239	LQNVDEEDDSYVLPITGESEDDNDEYIDQVYVYGEALSTV--MDNILSGFENEYDVIYKPL	1357
QY	1290	AGVYRSJKQJLENNVMTVNNVAVDKILNSRENKREKRNIALESJLIPYKJUTGSNNVAVDP	1349
Db	1358	AGVYRSJKQJLENNVMTVNNVAVDKILNSRENKREKRNIALESJLIPYKJUTGSNNVAVDP	1411
QY	1350	YKFLMKERDKFLSSYIYIKDSIDTQJINPANDVLYGYIKLISEKYSDDLSDISIKYI-----	1404

Dh	1418	FKLINSSEKNTLKSYYTKEISEVNOITKFAQESISYKEVLAKYDDLESIKKYIKEKE	14777
Qy	1405	-----NDKQSENEKYLPFLNINIEYLYKTVNDKIDLFVJHLEAKVLYNYE	14449
Dh	1478	KFPSSPTTPSPAKTDEOKRESEKFLPFLINIEETLYNNLVNKIDYLYLNTKAKINDCAYE	15373
Qy	1450	KSNEVAKTELYNTKYTDQKDLADFPKNNNVGJADJSTYNYNNINLLTKPLSGWEMEA	15099
Dh	1538	KDEAHVKTITJSDJAKAIDDKIDJLFKNNHNDDAKTKILINDTKKMDJGKLSGLGV-QNP	15566
Qy	1510	KTVSNLIDGNLQGMNLISQHCVKRQCPDQNSGCFPHLDERECCCLNYKQEGDKYEN	15699
Dh	1597	NTIISKLEGGFQDMLNISQHCVKRQCPDQNSGCFPHLDERECCCLNYKQEGDKYEN	15666
Qy	1570	PNPITCENENGGCADACTEDSDGSGMKKITCECTPDSYPLFDGIFGSSSNFLGISTLL	16239
Dh	1657	PNPITCENENGGCADACTEDSDGSGMKKITCECTPDSYPLFDGIFGSSSNFLGISTLL	17166
Qy	1630	ITMLILYSFI 1639	
Dh	1717	ITMLILYSFI 1726	

```

RESULT      5
MSPL_PLAFM  STANDARD:      PRT:  1701 AA.
ID   MSPL_PLAFM
AC   P08569;
DT   01-AUG-1998 (Rel. 08, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE   (PMMSA) (P190).
GN   MSP-1.
OS   Plasmodium falciparum (isolate mad20 / Papua New Guinea).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=70153;
[1]
SEQUENCE FROM N.A.
MEDLINE=88011243; PubMed=3079521;
RX   Tanabe K., Mackay M., Goman N., Scaife J.G.;
RT   "Allelic dimorphism in a surface antigen gene of the malaria parasite
RL   Plasmodium falciparum";
RN   J. Mol. Biol. 195:273-287(1987).
[2]
REVISONS TO 1403; 1569 AND 1629.
RA   Tanabe K.;
RL   Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE OF 1-115 FROM N.A.
MEDLINE=86136024; PubMed=3004972;
RX   Mackay M., Goman N., Bone N., Hyde J.E., Scaife J., Certa U.,
RA   Stunnenberg H., Bujard H.;
RT   "Polymorphism of the precursor for the major surface antigens of
RL   Plasmodium falciparum merozoites: studies at the genetic level.";
EMBO J. 4:3823-3829(1985).
-1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
(POTENTIAL).
-1- PWM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
-----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb.ch/announce/
CC   or send an email to license@isb-sib.ch).
-----
DR   EMBL: X05624; CAA29112.1; -.
DR   PIR: A26868; A26868
DR   PIR: B25120; B25120.

```

DR InterPro: IPR000561; -
DR Pfam: PF00008; EGF 1;
KW Melaria; Merozoite; Polyporein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 536 536 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 607 607 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 919 919 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 991 991 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AF6BA98 CRC64;

Query Match 60.23; Score 5069; DB 1; Length 1701;
Best Local Similarity 60.18; Pred. No. 6.3e-151;
Matches 1042; Conservative 226; Mismatches 339; Indels 126; Gaps 29;

QY 1 MKIIFFLCSFLFIINOCVTHESYOELVKLEALDAVLGYSLFOKEKMYLNEGSGT 60
DB 1 MKIIFFLCSFLFIINOCVTHESYOELVKLEALDAVLGYSLFOKEKMYLNEGSGT 60
QY 61 AVTTSTPGSGSVASGSGSVASGSGSVASGSGSVASGSGSGSRTPNSDSSDS 120
DB 61 AVTTSTPGSGSVASGSGSVASGSGSVASGSGSVASGSGSGSRTPNSDSSDS 114
QY 121 DAKSYADLKHVRNYLLTTELKLPOLFDLTNHLTLCDNHGKXYLDGEEELNELLYK 180
DB 115 NTKYADLKHVRNYLLTTELKLPOLFDLTNHLTLCDNHGKXYLDGEEELNELLYK 174
QY 181 LNFYFDLLRAKLVANDVCANDQIPFNKIRANEDLVKLKLVGYRKLPLDNKDNVGMED 240
DB 175 LNFYFDLLRAKLVANDVCANDQIPFNKIRANEDLVKLKLVGYRKLPLDNKDNVGMED 234
QY 241 YIKKNNKTITENINELIESKKTIDKNKNATKEEKKKLYQAQYDLSLYNQLSEBAHLIS 300
DB 235 YIKKNNKTITENINELIESKKTIDKNKNADBEKKKLYQAQYDLSLYNQLSEBAHLIS 294
QY 301 VLEKRIDTLKKNKIRKELDKINIK ---NPPRANGSNPTNLLD-KNKKIEHEKEIK 355
DB 295 VLEKRIDTLKKNKIRKELDKINIK ---NPPRANGSNPTNLLD-KNKKIEHEKEIK 351
QY 356 ELAKTIFNIDSLFTDPLELEYLREKKNNDISAKVE--TKRESTEPN-EYRNGVTPPLS 412
DB 352 ELAKTIFNIDSLFTDPLELEYLREKKNNDISAKVE--TKRESTEPN-EYRNGVTPPLS 411
QY 413 YNDINNAL---NELNSFGDLINPFDYTKEPSKNITYDN-ERKKPFINIEKEIKIEKKKI- 467
DB 412 LFTDIHNSLADNDKNSYDGLMNP-DTKEKINELIITDNKEKKIIFINNKKQIDIEENIN 470
QY 468 ---ESDKKSYEDSKSLNDITTEYKELKLNELYSKFNNDIDLNFEKMGKRSYKYEKL 524
DB 471 HTRKQNKRLLEDEKRS---KDYELLLEKFEYKMFNNNDKDVADKIFSRVAYNVEKK 526
QY 525 THNHTFASYSKHNLEKTLKALKYMEDYSLRNTVVEKELKYKYNLSKINENETEVEN 584
DB 527 RYNNKSESSNNYVNOYKLLKALSYLEDYSLRKISSEKDRNHYTLTGLEADKIKLLEE 586
QY 585 IKRDEQGLFEKK---TKDENKDPDEKILLEVSDIYKVOYVYLLMKNIDELKKTOLLKNV 641
DB 587 IKSEENKILKKNFGLTHSANAS---LEVSDIYKVOYVYLLMKNIDELKKTOLLKNV 642
QY 642 ELKHNHVPNSYQKQKQEPYLYLVKKEIDKLVMPKYESLINEKKNKITEGSDNS 701
DB 643 QLDSDIHVPNIYKQKQEPYLYLVKKEIDKLVMPKYESLINEKKNKITEGSDNS 695

QY 702 EPSTGEITGOATTKPGQAGSALSGDSVQAQAEQKA-----QPPVPVYPEAKAQ 754
DB 696 --ITQPLVASETTEDEGGSTHTL-SQSGETEVEETEVEETGHTTYYITLPPKEES 752
QY 755 VPTPAPVANNKTEN-----VSKLDYLEKLYEFLMTSYCHKYIIVSHSTAMEKILKQ 806
DB 753 APKEVKVYENSIEHKSDNSQALTKTYTKLDEFLKRSYCHYIIVLSNSMQKLEEV 812
QY 807 YKTRKEESKSLSCDDPLDLFNITONNIPWATSMFDSLNSLSQLPMEIYEKEMVCMYKL 866
DB 813 YNLTPKEEKELKSCDDPLDLFNITONNIPWATSMFDSLNSLSQLPMEIYEKEMVCMYKL 872
QY 867 KDNKIKKLEAKV-----STVYKLSLSSS-----MQPLSTPDQKREVS 908
DB 873 KEENHKKLEAKV-----STVYKLSLSSS-----MQPLSTPDQKREVS 932
QY 909 AND---DTSHS---TNLNSLKLENTLSGKKNKIYQEL-IGOKSENFYERILKDS 960
DB 933 SGPAVESHDPPLVLSLNDLKGIVSLNKGNTKYPNPLITSTTEKEKYENILKND 992
QY 961 TFYNESTPNYKSKADINSLNDESKRKKLEEDINKAKTILQSLFDLYNKKYKLERLFD 1020
DB 993 TYFNDDIKQFVKSNSKYITGLT-ETOKNALNDEIKLKLDLQSLFDLYNKKYKLERLFD 1051
QY 1021 KKTIVGKYKMOIKKLTLLKEQLESKLSLNNPKRYVLONFSVFYKKEAEIAETENTLEN 1080
DB 1052 KKKELGQDKMOIKKLTLLKEQLESKLSLNNPKRYVLONFSVFYKKEAEIAETENTLEN 1111
QY 1081 TKILLKHYKGLVKNYNGESSPLKTLSESIQTEDNVASLENFKVLSKEGKLNKLNLEK 1140
DB 1112 TKILLKHYKGLVKNYNGESSPLKTLSESIQTEDNVANLEKFRALSKIDGLNNDLHKG 1171
QY 1141 KKLTSLSGLHLLAELEKVKYKNNKNTGNSPENNNTYNNALSEYKFLPBGTYAAYVS 1200
DB 1172 KKLTSLSGLHLLAELEKVKYKNNKNTGNSPENNNTYNNALSEYKFLPBGTYAAYVS 1230
QY 1201 E-----SGSDTLEQSOOPKPPASTHYGAESNTTTSQNDVEDVYIPIF 1246
DB 1231 PPGQDYVTPSPLSYAVSGSSSGSTKRETFQIPPTSGSLTLEQYQVQYQNDDEEDSLVPIF 1290
QY 1247 GESEEDYDGLQVVTGAVTPSVIDNLSKLENEYLYLPLAGVYRSKLEQLENNVT 1306
DB 1291 GESEEDYDGLQVVTGAVTPSVIDNLSKLENEYLYLPLAGVYRSKLEQLENNVT 1349
QY 1307 FNVAVKDIINSRFRKREKRVNLESDILPKYKDLTSSNYVAKPKFLNKKERKDFLSSYN 1366
DB 1350 FNVAVKDIINSRFRKREKRVNLESDILPKYKDLTSSNYVAKPKFLNKKERKDFLSSYN 1409
QY 1367 YIKSDIDINFAVDVGYKRIISEKYSKSDLSIKYI-----ND 1406
DB 1410 YIKSEVEDIKFAQEGSIYEKYLAKTKDDLESTIKYIKEKEKFPSSPTTPSPAKTD 1469
QY 1407 KQGENEKYLPFLNIETLYKYTVNDKIDLEFVTHLEAKVANTYKESNVEYKIKELNYKTI 1466
DB 1470 EKKESKFLPLPLNIEFLYNNLVNKKIDYILNLKAKINDCVAEDEAHVKTIKLSDLKAI 1529
QY 1467 QDKLADFRKNNNFGIADLSTVYNNHNLTKFLSTGAVFENLAKTVLSNLDGLMGLN 1526
DB 1530 QDKLADFRKNNNFGIADLSTVYNNHNLTKFLSTGAVFENLAKTVLSNLDGLMGLN 1588
QY 1527 ISOHQYKQKOPNGSGCFRHLDERECCCLNLYKQEBDKCVENPNPCNENNGGCDADAK 1566
DB 1589 ISOHQYKQKOPNGSGCFRHLDERECCCLNLYKQEBDKCVENPNPCNENNGGCDADAT 1648
QY 1587 CTEBDSGNGKKTICETKPDSPYLPFDGIFCSSNFGISFLILMLILYSFI 1639
DB 1649 CTEBDSGNGKKTICETKPDSPYLPFDGIFCSSNFGISFLILMLILYSFI 1701

RESULT 6
MSPL_PLAFF STANDARD; PRT; 1701 AA.
ID MSPL_PLAFF

AC P13819;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
 DE (PMSA).
 GN MSP-1.
 OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID:5837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8142999; PubMed=2449612;
 RA Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,
 RA Brown G.V., Anders R.F., Kemp D.J.
 RT Variation in the precursor to the major merozoite surface antigens
 RT of Plasmodium falciparum.
 RL Mol. Biochem. Parasitol. 27:291-302(1988).
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POTENTIAL).
 CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
 CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M19143; AAA29653.1;
 DR PIR: A54498; A54498.
 DR InterPro: IPR000561;
 DR Pfam: PF00008; EGF. 1.
 KW Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 536 536 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 607 607 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 919 919 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 991 991 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 1701 AA: 193719 MW: 3920875E73D38552 CAC64;

Query Match 60.1%; Score 5065; DB 1; Length 1701;
 Best Local Similarity 60.1%; Pred. No. 8.3e-151;
 Matches 1042; Conservative 227; Mismatches 338; Indels 126; Gaps 29;

QY 1 MKIIIFLCSTFLPINTOCYTHESYGLVKKLEALDAVLGTGSLFQEKVNLNEGSGT 60
 DB 1 MKIIIFLCSTFLPINTOCYTHESYGLVKKLEALDAVLGTGSLFQEKVNLNEGSGT 60
 QY 61 AVTTSPGSGSVASGSGSVASGSGSVASGSGSVASGSGSRRTNPDSNDS 120
 DB 61 AVTTSPGSGSVASGSGSVASGSGSVASGSGSVASGSGSRRTNPDSNDS 120
 QY 121 DASYADLKRVRNYLTLEIKYQPLFDTLNLHLLCDNHGKYYLIDGEEINELLYK 180
 DB 115 NMTYADLKRVRNYLTLEIKYQPLFDTLNLHLLCDNHGKYYLIDGEEINELLYK 180
 QY 181 LNFYDLRAKLVNDVCAVCOIPFNLIKIRANELDLVKKLVFGYRKPLDNIKDNVGMED 240

DB 175 LNFYDLRAKLVNDVCAVCOIPFNLIKIRANELDLVKKLVFGYRKPLDNIKDNVGMED 234
 QY 241 YIKNKKTIEINILIEESKTTIDKNNATKEEKKKLYQAQDLSYNNQLEAHNLIS 300
 DB 235 YIKNKKTIANINILIEGSKTTIDKNNADNEEGKKLYQAQVNLFIYNNQLOEAHNLIS 294
 QY 301 VLEKRIPTLKKNNIKELDLKINEK---NPPANGNPNTLLD-KNKKIEHEKEIK 355
 DB 295 VLEKRIPTLKKNNIKELDLKIKTDAENP---TGSKRNPPLPENKKEVEGHEKEIK 351
 QY 356 EIAKTIKENIDSLPTDPLELEYLREKNNIDISAKVE--TKESTEPN-EYPNGVTPYLS 412
 DB 352 EIAKTIKENIDSLPTDPLELEYLREKNNKVDYPSQDPFKSQVQIKVPNGIVPLP 411
 QY 413 YNDINNAL---NELNSGDLINPDYTKPEFSKNTYTDN-EKKKINIEKIKIEKKTI- 467
 DB 412 LTTDHNLSIADNDKNSYGDLMNP-DTEKEINEKLTTPNKEKRTIFINNKKQIDLEEKIN 470
 QY 468 ---ESDKSYEDRSKSLNDITREYKELINLEYDSKFNNNIDLTNFKMMGKRYSYKEKL 524
 DB 471 HTKQNKRLLEDYERK---KDIIELEKFEYKKNFNNNDKVDYVDKIFSAITYYNEKQ 526
 QY 525 THNTFASYSKNNLEKLTAKLYMEDSLRNIVREKELKYKNLSIKTENEIETLVEN 584
 DB 527 RYNNKFPSSNNSVYNOKLLKALSYLEDYSLRKGISKDPNNHYTTLTGLEADIKLTKEE 586
 QY 585 IKKNEEDLFEEK---TKDENKPEKLTLEVSQVQVQVYVLMNKKIDELKQIOLIKV 641
 DB 587 IKSENKILEKKNFKLHNSANAS---LEVSDIVKLOVOVLIRKIFEDLRKTELPLFA 642
 QY 642 ELKNNIHVPNSYKQENQOEYVTLVKKKEIDKLKVPKYESLINNEKNNIKTEGQSDNS 701
 DB 643 QLKDSIHVPNTIKQKNPPEYVTLVKKKEIDKLKVPKYESLINNEKNNIKTEGQSDNS 701
 QY 702 EPSTGEITQATTPKQQAQSALEGDSVOAQAQKQA-----OPVPVYVPEAKAQ 754
 DB 696 --ITQPLVAASETTEDEGSHSTHTL-SQSGEVEVEETEVEETVGHHTYVITLPPKEES 752
 QY 755 VPPRPVAVNNKTE-----VSKLDYLEKLYEFLNTSTYCHKLYTVSVSTNNKTLKQ 806
 DB 753 APEKVVAVNSIEKSDNSQALTKYVLRKLEFKYSTYCHKLYTVSVSMDOKLLEV 812
 QY 807 YKITEEESKLSQDPLDLFNQNNIPVMYSMEDSLNLSLSQLFMEYKKEWVCNLYKL 866
 DB 813 YNLPPEENELKSDPDLFLFNQNNIPVMYSMDSLNLSLSQLFMEYKKEWVCNLYKL 866
 QY 867 KDNKIKNLLEBAKV-----STYKYLSSSS-----MQLSLTPQDKPEVS 908
 DB 873 KEENHIKLLLEBOKLIGTSTSPGNTVTVAQSATHNSQOQNASSTNONGVAVS 932
 QY 909 AND---DTSHS---TNLNSLKLFEMLISLGNKNKIYQEL-TGQKSEMFYEKTLKSD 960
 DB 933 SGPAVVEESHDPLTVLSISDLKGIYSLNIGNKTKVNPDLTSTYEMEFYEYILKNND 992
 QY 961 TFYNESFTNFVSKADINSINDSKRRKLEEDINKLTKTQLSFDLYNNKYLKLERLFD 1020
 DB 993 TYENDIDKQGVKSKYITGLT-ETQKNALNDELKIKKIDQLSFDLYNNKYLKLERLFD 1051
 QY 1021 KKTIVGKYYKQKIKKYLKQLESKLSLNNPKHVLONSVFVFNKKKEAIAETENTLEN 1080
 DB 1052 KKEELGQDKQKIKKYLKQLESKLSLNNPKHVLONSVFVFNKKKEAIAETENTLEN 1111
 QY 1081 TKLILKHYKGLVYVYNGESSPLTLESSTIOTDNYVASLENFVLSKLEBKLDONLLEK 1140
 DB 1112 TKLILKHYKGLVYVYNGESSPLTLESSTIOTDNYVASLENFVLSKLEBKLDONLLEK 1171
 QY 1141 KKLVSLSGSLHLLIAELKEVYIKNNKNTGNSPSENNDVNNALSYKFLPEGTVAVTYS 1200
 DB 1172 KKLVSLSGSLHLLIAELKEVYIKNNKNTGNSPSENNDVNNALSYKFLPEGTVAVTYS 1230
 QY 1201 E-----SGSDTLEQSOPKRPASTVHGAESNTITTSQNDVDEVDVITVPIF 1246


```

Db 639 SHHPNITKPKQNKDEPYLYLVKKEVDLKEFTPKVKMDLKEQAVLSS-----IT 689
Qy 706 EGETTGQATKPGQAGSALGSDVQAQAOQOK-----AQPVPVPVPEAKAAY 755
Db 690 QPVAASEETEDGHSHHTLSQSGEETVEETETETVGHVTTTTLTPKPKAVVENSIEH 749
Qy 756 PTPAPVNNKTKENSKLDYLEKLEFLNTSYCHKYLIVSHSTNNKTLKQYKTKREES 815
Db 750 KS-----NDNSQALTKYVLYLKLDELFLTKYCHKYLIVSSNDQKILEVYNLT-PEEN 803
Qy 816 KLSGCDPLDLFNLIQNNIPVWYSMEFSLNLSQLFMEIYKEMVCMYLKLDKDKLNL 875
Db 804 ELKSCDRDLFLFNQNNIPVWYSMEFSLNLSQLFMEIYKEMVCMYLKLDKDKLNL 863
Qy 876 LEEAKV-----STKTKLSSSS-----MPLSLTPQDKREYAND---DTS 914
Db 864 LEEPKQITGTSSTSSPGNTVYTAQSAHSNSQNOQSNASTNTQNGVAVSSGPAVEES 923
Qy 915 HS-----TNLNSLKLFLNLSLGNKNKIYQEL-IGKSSSEYFKILKDSPTFNESFTN 969
Db 924 HDPPLTVLSTISNDLKGIVSLNLGNKTKVPNPLITSTEMEFTENILKIMPIFNDDIKQ 983
Qy 970 FVSKADINSLNDESKRKKLEEDINKLKTLOLSPDLNRYKYLELFDKRTVKYK 1029
Db 984 FVKSNSVYITGLT-ETQKNALNDEIKKLKDTLOLSPDLNRYKYLELFDKRTVKYK 1042
Qy 1030 MOIKLLFLKLEESKINSUNPKHVLQNFSEVFNNKKKEALIEFNTLTKLILKYY 1089
Db 1043 MOIKLLFLKLEESKINSUNPKHVLQNFSEVFNNKKKEALIEFNTLTKLILKYY 1102
Qy 1090 GLVYKNGESSPLTKLSEESIQEDENYASLENFVYLSKLECKLKDNLLEKKLSYLSG 1149
Db 1103 GLVYKNGESSPLTKLSEESIQEDENYANLEKFPVLSKIDKLNHLGKKLSYLSG 1162
Qy 1150 LHLIAELKEVYKKNKNTGNSPENNDVNNALSYKFLPEGDVATVSE----- 1201
Db 1163 LHLIAELKEVYKKNKNTGNSPENNDVNNALSYKFLPEGDVATVSE----- 1221
Qy 1202 -----SGSDPLEGOSOPKRPASTVGAESNTITTSQNVDDVDVYIPIGSEEDDD 1255
Db 1222 PLASRVGSGSSSTKEEQIPIPSGSLTELQOYVOLQNTDEEDDSIVLPIGSEEDNDY 1281
Qy 1256 LGQVVTGEAVTPSYIDNLSKIEENEYVLYLPLAGVYRSLLKQLENNVMTFNVNVDIL 1315
Db 1282 LDQVVTGEAVISVT-MDNILSGFENEDYIYLLPLAGVYRSLLKQLENNVMTFNVNVDIL 1340
Qy 1316 NSRPKNKRNKRVNLESDELIPKDLTSSNYVVKDPYKFLNKKRKFSLSYIKDSIDTD 1375
Db 1341 NSRLKRRKRYPLDVLESIDMOKHISISNEYIIEESFKLINSQOKNTLLKSYKIKESVEND 1400
Qy 1376 INFANDVLYGYKILSEKYSKDLSDISKYI-----NDQGENEKLPL 1416
Db 1401 IKFQOEGISYIEKLAAYKDDLESITKVIKEKEFPSSPPTPPSPAKTDQCKESKFLP 1460
Qy 1417 FLANNIETLYKTVNDKIDLEVLHLEAKVLYNTYKESNVEVRIKELNYLKTIDOKLADFEKN 1476
Db 1461 FLTIETLYNNLVKIDYIINLAKINDCVERDEAHVKITKISDLKALIDOKIDLEFRNP 1520
Qy 1477 NNEVGIADLSDIYVHNHNLKFLSTGWFENLAKTVLSNLDGMLQGMILSIHQCYAKQ 1536
Db 1521 YDFEAIKRLINDTDKRMKGLSTGLV-QNFPNTIISKLITGEGFQDMLNISQHCYAKQ 1579
Qy 1537 CPQNSGCFRHLDEEECKCLINTYKQEGDKCYENPPTCNENNGCCDADAKTEEDSGSNG 1596
Db 1580 CPQNSGCFRHLDEEECKCLINTYKQEGDKCYENPPTCNENNGCCDADAKTEEDSGSNG 1639
Qy 1597 KKINCECTKPDSPYPLFDGIFCCSSNPLGISPLILMLILYSFI 1639
Db 1640 KKITCECTKPDSPYPLFDGIFCCSSNPLGISPLILMLILYSFI 1682

```

```

MSPL PLAYO
ID MSPL PLAYO STANDARD; PRT; 1772 AA.
AC p13828;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMSA) (230 KDA).
GN MSP-1.
OS Plasmodium berghei yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=5862;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90205979; PubMed=2320061;
RA Lewis A.P.;
RT "Sequence analysis upstream of the gene encoding the precursor to the
  major merozoite surface antigens of Plasmodium yoelii."
RL Mol. Biochem. Parasitol. 39:285-288(1990).
RN [2]
RP SEQUENCE OF 1093-1772 FROM N.A.
RX STRAIN=17XL;
RX MEDLINE=86124889; PubMed=2448778;
RA Burns J.M., Jr., Daly T.M., Vaidya A.B., Long C.A.;
RT "The 3' portion of the gene for a Plasmodium yoelii merozoite surface
  antigen encodes the epitope recognized by a protective monoclonal
  antibody."
RT Proc. Natl. Acad. Sci. U.S.A. 85:602-606(1988).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
  (POTENTIAL).
CC -1- PPM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
  KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
  MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----
CC EMBL: J03612; AAA29702.1; -.
CC EMBL: J04668; AAA29702.1; -.
CC PIR: A28121; A28121.
CC PIR: A45532; A45532.
KW Malaria; Merozoite; Polyporein; Repeat; Signal; Glycoprotein;
  Transmembrane; GPI-anchor.
FT SIGNAL 1 18
FT CHAIN 19 1772
FT CARBOHYD 54 406
FT CARBOHYD 646 646
FT CARBOHYD 829 829
FT CARBOHYD 1018 1018
FT CARBOHYD 1090 1090
FT CARBOHYD 1408 1408
FT CARBOHYD 1446 1446
FT CARBOHYD 1541 1541
FT CARBOHYD 1629 1629
FT CARBOHYD 1680 1680
FT CONFLICT 1521 1521
FT SEQUENCE 1772 AA; 197230 MW; 9A6291658B0F45D CRC64;

```

Query Match 30.3%; Score 2548.5; DB 1; Length 1772;
 Best Local Similarity 32.4%; Pred. No. 1.3e-72;
 Matches 609; Conservativity 331; Mismatches 589; Indels 349; Gaps 40;

```

Qy 1 MKITFPLCSPLFIINQCVTHESYQELVKKLEDAVLGTSLFQKRYVNTGEGSGT 60
Db 1 MKVIGLGFSEVFVFAIKCKSETIEVYNDLQIKLEKLESLSVDGELTFQKSOVILN----- 54

```



```

OY 61 AVTTSTPGSKSVASGGSGSVASGGSVASGGSVASGGSGNRRTPNSDSSDS 120
Db 55 -----ATOPETIDPF 65
OY 121 DAKSVADLKHVRVNYLLTIRKELYPOLFDLTNNHMLTLCNDINHGRYLLIDGVEEINELLYK 180
Db 66 TNNHFA---QOVQDFYKKEGLGFTQETELVNLKALPNNYGVYLLIESKEEFGMLHA 122
OY 181 LNFYEDLLRAKLVNCANDYCOIPEFNKIRANELDVLKLLFGYRKPLDNKDNVGMED 240
Db 123 INFYEDVLRDLNMCANNICEIPEHLKISEETEMLKVLLGTRKPLENIQDDIEKLEI 182
OY 241 YIKNNKKTENINELIESKKT-----DKNNKATKEEEKKLYOAOYDLSYNNKQ 291
Db 183 YIERNKETVALNMLIAEETKIQPEGNECDNDASCDSDKNKKKPIYQANNYVTFYKKQ 242
OY 292 LEEHANLISYLERKIDPLKKNENKELLDKINEIKNPP-----PANSG 334
Db 243 LAETOKVVEVLEKREVSLLKNDKAIKPLMOQIEVLNAAPVVTAETQIVTGSGSSTPGSGG 302
OY 335 N-----TPN-----TLLDKN 344
Db 303 SSASGTSSSGASAGTVEQANIVASVTVPSVQONGEASTNPOTAOVYPVTLTLEKQ 362
OY 345 KRIEHEKEIKETIKTFNIDSLFTDPLEEYLYREKNK---NIDISA--KVEFKESTE 399
Db 363 KKIINGVIAQIKETIKTFNIDSLFTDPLEEYLYREKNK---NIDISA--KVEFKESTE 399
OY 400 P--NEYNGVTVPLSYNDI-----NALNELNSTGDLINPVDYKREPSKIYTDNERKKF 452
Db 423 PLTRYNGVTVPLSYNDI-----NALNELNSTGDLINPVDYKREPSKIYTDNERKKF 452
OY 453 INEIKETIKETIKETIKESDERSKSLDITKEVEKELNEIYDSKRNNDITLNEPKM 512
Db 480 IKAIKKIKETIKETIKETIKESDERSKSLDITKEVEKELNEIYDSKRNNDITLNEPKM 512
OY 513 MGKRYSVKVELTHTNFASYENSKHNEKLTAKAKMEDYSLNIVYKELKAYKNLIS 572
Db 540 KTRKDEYMTK--TELNT--CEGNTKELINKNOLANTLODSLKDIISNIEIEFNSKKK 597
OY 573 KIENEIETLVENIKKDEQOLFEEKITQDENKPDKEILEVSDIVAVQOVKVLNKKIDELK 632
Db 598 ELQVNIIRLEAVQAKONVIA-----SKDVPPLSTLVEIQLOKSLTTLQIOEOLN 646
OY 633 KTOIILKNELKHNHIVPNYSKOENKOEPPYLLIYKKEIDOKLAFMKRVESLNEEKKNI 692
Db 647 KTEVSLKAQOLKDLKLYPKTYGNEGKPEPYLLIAVKKVEVDLAQIFPIESIMAKKERM 706
OY 693 K-----TEGQSDNSEPSTEGEITGOATTKPGQAGSA-----LEG 727
Db 707 EOGPAITGESEEPVSGPSAESTDRSSTSSSSSSSTPAALAESSATILPEAPAPAEA 766
OY 728 DSVQAOQEOKQOPVPPVPEAKAOPVPPAPVNNKTEVNSKLDYLEKLYEFLNTSYI 787
Db 767 ASPSTASEETITPTTQETOPQSOAASSTPAKPV-----MTKLYLEKLOKLFVSYS 820
OY 788 CHKYILVSHSTMEKILKOKITKEES--KLSCDPLDLFNIONNIPVYMSFEDSLN 845
Db 821 CHKYVLQNSTINKDALSKALISEEKIRLTKCSELDDLALIONMPTMYSLESTIV 880
OY 846 SLQOLFMEIYEKEMWCLYKLD--NDKIKMLLEBA-----879
Db 881 GLQNIYVELYKEMHYIKLKDENPSIKSLVYKAGVIEPEVAAVPPVPAATEQOQOQ 940
OY 880 --KVVTSVATLSSSQPLSLPQDKPEVSANDDTSHS-----916
Db 941 ATPDVQSDAPAPDSVQOPETPVTTSTPTSTPEASSAPAGECTPSGEGAGASCTEGATA 1000
OY 917 -----TNLNSLKLFEENILS-----LCKNKIYBELI 943
Db 1001 SNATPAGTASGSAASNASTTSVTPPAAAAPVSTSTPAPAPPPAASQSGNPDGICIR 1060
OY 944 GOKSSE-----NFEYKILKDSDFYNESTFNVFSKADINDSLNDESKRKLTEE 992

```

```

Db 1061 SRASEEDMPADDFELDLNLYKSYLQOJDG--NNTFEINIFKSKKELIKALTPD-KVNOPLYL 1118
OY 993 DINKIKTLQSLPDLNLYKSYLQOJDG--NNTFEINIFKSKKELIKALTPD-KVNOPLYL 1052
Db 1119 EIAHLKELSEHYDRYSYKTLERLKNHBOIOLJTNQOJLISILKARLLKRRQOTLNGV 1178
OY 1053 KHAVLONFVFNNKKKEAFIAETENTLENTKILKHYGLVYVYNGESSPLKTLSESIQT 1112
Db 1179 FYILNGVNFNNKREAEKQYVNAKNTDMLLYYKARTYFTSEAPVPLTSLKASIDR 1238
OY 1113 EDNYSLENFVLSLEKTLNLEKRLNLEKRLNLEKRLNLEKRLNLEKRLNLEKRLN 1172
Db 1239 ESNVLKIEKFRAYSRLERLKNKINILKERSISYSGGLAHVEEFKELIKODYTKKNP 1298
OY 1173 ENMTDVNNALESYKFLPEGTDAVTVSSEGSJLLEQSOPEKPASTHVAESNTI----- 1227
Db 1299 DNAPEVTNAPEQYKELLIPKGVYST--PAVAVTTTLADAPATPEGAVPAGVAVP 1357
OY 1228 -----TTSQNDDEVDVLIIVPIGESFEDDDLGQVYTGAVVPSVIDNLSK 1276
Db 1358 GAVPAGVPGSGTDRVAGSSVDD-----NED--DDIYQIASGQSEDAPEKD-ILSE 1405
OY 1277 IENEYEVLYLKLPLAGVYRSLKOLENNVMEFNVAVKIILNSFRKRENFKNVLESDLIPY 1336
Db 1406 FTNESLIVYTRKLSQYKSLKHLKREFTIKEDMTGNLNNKSKRNDPFLVLSHEDLF 1465
OY 1337 KDLTSSNVVADPKYFLNKKERKDFLSSYNYIKDSIDTIDFANDVLYGYYKILSEKXSD 1336
Db 1466 KDLSTNKVYINRNPQLDNDKDKQOIVLKYATKGINEDIETTTGDKIFKNKVELYNQO 1525
OY 1397 LDSIKKYI-----ND--KQENKRYLPLNINIEYLVKTVNDKIDLEVIHLEAVLWYTY 1448
Db 1526 LAAYKEQATIEATNTDNNKKEKKYIPLEEDLGLVETVYQAEVSEELONLNDYKN 1585
OY 1449 EKSNVKIKELNLYLKTIDKLADFE---KKNNFVGIADLSTYNNHNLITKFLSTGMV 1504
Db 1586 EKAEFELITKNLEKLYQIDEDLDEFVHAENKNIHIAIA-----LNNLNKGGIV 1634
OY 1505 FENLAKTVLSNLDGNIQGM--LNIS--OHOCV--KQCPONSQCFRHLDERECKCLNLYK 1560
Db 1635 GEGESKRIKML--NNDGMDLGVDPRKHCVDPRDIPKNAGCFRDNGTEEMCLGKY 1692
OY 1561 Q-EGDKVCENPNPCNNNGCCDADACTEDSGSNKKITCECTKXDSVPLFGICSS 1619
Db 1693 KQEGNTCVENNPPCDLNNGCCDPTASQNAESTENSKITICTCEPTPAVAYGVCCSS 1752
OY 1620 SNFLGISFLILMLILYS 1637
Db 1753 SSFNGSLILITLITLIVFN 1770

```

```

RESULT 9
MSPL_PLAFD
ID MSPL_PLAFD STANDARD; PRT: 233 AA.
AC P13827:
DF 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMMSA) (GLYCOPROTEIN 185) (GP185) (FRAGMENT).
OS MSP-1.
SN Plasmodium falciparum (isolate CDC / Honduras).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
RX NCBI_TaxID=5836;
RN [1]
RP SEQUENCE FROM N.A.
RM MEDLINE=87106855; PubMed=3542719;
RA Howard R.F., Ardeshir F., Reese R.T.;
RT "conservation and antigenicity of N-terminal sequences of GP185 from
RL different Plasmodium falciparum isolates.";
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR

```

(POTENTIAL).
-1- P1M: MEROZOITE SURFACE ANTIGEN WHICH ARE THE MAJOR SURFACE ANTIGENS OF
KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M14632; AAB59227.1; -
DR PIR: A25814; A25814.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19
FT CHAIN 20 >233 POTENTIAL.
FT CARBOHYD 104 104 MEROZOITE SURFACE PROTEIN 1.
FT NON_TER 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 233 AA; 26050 MW; 1DE4D31ACAE965 CRC64;

Query Match 14.1%; Score 1187; DB 1; Length 233;
Best Local Similarity 95.1%; Pred. No. 2.8e-31;
Matches 232; Conservative 0; Mismatches 0; Indels 12; Gaps 1;

QY 1 MKIIFLCSEFFIINOCVTHESYOELVKKLEADAVLTGYSLFQEKKNVLENGTSGT 60
DQ 1 MKIIFLCSEFFIINOCVTHESYOELVKKLEADAVLTGYSLFQEKKNVLENGTSGT 60
QY 61 AVTSTPGSGSVASGSGSVASGSGSVASGSGSGNRRPNPSPNSDS 120
DQ 61 AVTSTPGSGSVASGSGSVASGSGSVASGSGSGNRRPNPSPNSDS 108
QY 121 DAKSYADLKHVRNYLLTIKELKYPOLFDLTNNHMLTLCDNHGFYLLIDGYEINELLYK 180
DQ 121 DAKSYADLKHVRNYLLTIKELKYPOLFDLTNNHMLTLCDNHGFYLLIDGYEINELLYK 168
QY 169 LNFYFDLIRAKLNDVANCYCOIPFNKIRANEDLVKLVFGYRKKPLDNKIDVGMED 228
DQ 169 LNFYFDLIRAKLNDVANCYCOIPFNKIRANEDLVKLVFGYRKKPLDNKIDVGMED 228
QY 241 YIKK 244
DQ 241 YIKK 232

RESULT 10
MSP1_PLAFN STANDARD; PRT; 281 AA.
ID MSP1_PLAFN
AC P13820;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (EMMSA) (FRAGMENT).
GN MSP-1.
OS Plasmodium falciparum (isolate NF7 / Ghana).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=5842;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=88142999; PubMed=2449612;
RA Peterson M.G., Coppe R.L., McIntyre P., Langford C.J., Woodrow G.,
RA Brown G.V., Anders R.F., Kemp D.J.;
RT "Variation in the precursor to the major merozoite surface antigens
RT of Plasmodium falciparum."
RL Mol. Biochem. Parasitol. 27:291-302(1988).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -1- MISCELLANEOUS: MEROZOITE SURFACE ANTIGEN WHICH ARE THE MAJOR SURFACE ANTIGENS OF

83K, 42K & 19K ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M19144; AAA26354.1; -
DR PIR: B54498; B54498.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19
FT CHAIN 20 >281 POTENTIAL.
FT CARBOHYD 109 109 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 281 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 281 AA; 31226 MW; 6150B76E0F21B299 CRC64;

Query Match 13.0%; Score 1096.5; DB 1; Length 281;
Best Local Similarity 77.7%; Pred. No. 2.2e-28;
Matches 227; Conservative 8; Mismatches 32; Indels 25; Gaps 5;

QY 1 MKIIFLCSEFFIINOCVTHESYOELVKKLEADAVLTGYSLFQEKKNVLENGTSGT 60
DQ 1 MKIIFLCSEFFIINOCVTHESYOELVKKLEADAVLTGYSLFQEKKNVLENGTSGT 56
QY 61 AVTSTPGSGSVASGSGSVASGSGSVASGSGSGN-SRRTN----- 112
DQ 57 EITR-----KG-ASAQSGASQSGASQSGASQSGSGTSGSGSGTSPSSRSITLRSN 109
QY 113-----PSDSSDSDANSYADLKHVRNYLLTIKELKYPOLFDLTNNHMLTLCDNHGFY 166
DQ 110 TSSGASPPADASDSDANSYADLKHVRNYLLTIKELKYPESLDLPNNHMLTLCDNHGFY 169
QY 167 LIDGYEINELLYKLNRYFDLIRAKLNDVANCYCOIPFNKIRANEDLVKLVFGYR 226
DQ 170 LIDGYEINELLYKLNRYFDLIRAKLNDVANCYCOIPFNKIRANEDLVKLVFGYR 229
QY 227 PLDNKIDVGMEDYIKKNTINELIESKKTIDKNATKEEKKL 278
DQ 230 PLDNKIDVGMEDYIKKNTINELIESKKTIDKNATKEEKKL 281

RESULT 11
YD86_SCHPO STANDARD; PRT; 1957 AA.
ID YD86_SCHPO
AC Q1041;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 222.8 KDA PROTEIN C1F3.06C IN CHROMOSOME 1.
GN SPAC1F3.06C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_Taxid=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA STFAIN=972;
RA Connor R., Church C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: Z70690; CA94624.1; -
DR Hypothetical protein.
KW SEQUENCE 1957 AA; 222785 MW; 3f480CA06171D9DA CRC64;

Query Match 6.0% Score 501.5; DB 1; Length 1957;
Best Local Similarity 22.2%; Pred. No. 5.4e-09;
Matches 352; Conservative 253; Mismatches 617; Indels 361; Gaps 67;

QY 105 SGNSTRTPNSDDSDAKSYADLKHVRNRYLLTIKELQYFQDLTNHMLTLCNINCF 164
D 105 SGNSTRTPNSDDSDAKSYADLKHVRNRYLLTIKELQYFQDLTNHMLTLCNINCF 164
D 162 SSNNKDKMTSSVTTLTSEDVSYFOKLTN-MESNFKOSEADLSQDLTVTEKLK 220
QY 165 KYLDGEIEMELLYKLFYDLRAKLVNDVANDCOIPENLKRANLQVLYKLVGY 224
D 221 E-KDEKIKEDY-----SIIKASIAEQASN-----KSLRGEORLEKLIVSS 262
QY 225 RPLDNIKDNVGMEDYIKKNTIENINELIESKTIIDKNKNTKEEKKLYQAYD 284
D 263 NKIVSTLE---QTEENLRACEKTLQ---EKLE-----KCAINDESKLEELKHN 306
QY 285 LSIYNKOLEAHNLSVLEKRI-----DTLK-KENIKELLDKINEIKNPPANS 333
D 307 VANSADAIVHDKLIEDLSTRISEFDNLKSRDILSIKNELEKL----- 352
QY 334 GNPNTLID--KAKKEHEKEIKELAKTI-----KFNIDSLTFDPLEEYLRKN 383
D 353 RNTIGSLKDSRTSLSOLEEVEVELESNRTYHSQLTDAESKLSFEQENKSLGSDIEYQ 412
QY 384 KNIDISAKVERKESTEPREYNGVLYPLSYNDINNALNELSPDGL-----INPDYTERP 439
D 413 NNLSSKDKMAVQVSSOLEEANSIAH-----ATGKLAELINSEDPONKATKDEKIEOD 466
QY 440 SKNIYTDNERKKTINELIKIKIEKK-----IESDKSYEDSKSLNDITKEY- 488
D 467 LRACLNSS-----NELKEKSLIDKKQOELNLRQEQKQKYSSESQSSLOSLQRIIL 521
QY 489 -EKLINELYSKFN-----NIDLTFEKKMGKRYSTYVEK-----LTHNHTFASYSKRN 539
D 522 NEKKRHEVEYSELWELGLOTEIENSELSSQSLTLAEKEAVALATNNESESNSIOT 581
QY 540 L-----EKLTFALKYMEDYSLRNIVEKELKYKNLSIKINELIETLVENIKKEOD- 592
D 582 LCNAFQEKLAQVMOLENEQNESSLDTSFKKLNESHOELNHNHTYKOLKDTSSKLOQ 641
QY 593 -----FEKK---ITKDEKPRDEKILEVSDIVKVOYKVLMMKIDELKKTOLIKNVE 642
D 642 LQLEBRANPEQESTLSDENNDRFLKLKLEESKSLTK---QEDVDSLEKNIQTIKE-D 697
QY 643 LKHNIHPNASKQKQKQEPYLLIYKKEIDKLYKMPKVESLINDKKNIKTEGOSD--N 700
D 698 LRKSEALREFKLEAKN-----LREVIDNLK--GKHETL--EAQRNDHASSLSDAKN 745
QY 701 SEPSTEGEITQATTRGQAGSALLEGDSVQAQOEOKAOPVYPRVPEAKQAQVPPRA 760
D 746 TNALISSELT-----KSESDVKRLTANVETLTODSKAM----- 778
QY 761 PVNNKTENVS---DYLEKLYEFLNTSYICHKYLISHTSTNNEKILKOYKITEESKLT-SS 819
D 779 -KQSFSTLVNSYQISLNIYHELRRDH-----VNMOSQNTTL-----ESSKLTKD 823
QY 820 CDPL-----DLFNIQV-----NIPVYSFDSLNSLSLOLFEIYEKENVLCULYKID 868
D 824 CENITQONMTLIDVOKLMHKNVQESKVSLEKLVNGLS--LDLKLIRSSLNV-AISD 879
QY 869 NDKIKNLLEAKKYSTSVKTLSSSSMQPLSLTPQDKPEVSANDTQSH--STNNLNSKLTF 926
D 880 NDQILITQALAEISKYDSLEQSSAQLNSGLKSLAEKQOLLTENEELHRLDKLTGKAKIE 939
QY 927 ENILS-LGKKNKIYQELIGOKSSSENFYEKILKDSDFEYNESFTNFKVSKADINDLANDES 985

D 940 ESKSDLCKKTLIARQEEISLNKEENMSQ-----SQATIS-VKSKLDE-----TIS 983
QY 986 KKKLEEDINKKKTLOLSPDLVYKRYKLERLFDKKYTVK---YKQIKKLTLEKQ 1041
D 984 KSKLEADIEHLKKNKVSVEVERNALLASNRLMDLKNKGNENIASLOTETIEKRAEND 1043
QY 1042 LESKLNLSNPKHVLQNSVFENKKKEKAEIATETENTLNTKILKHYGLVKKYNGESSP 1101
D 1044 LQSKLSVSVSEYENLLISQTNKSLIEDKTNQKRYEKNVQGLKLEDKQORVLEELTSK 1103
QY 1102 LKTLSEESIQTEDVYASL-ENFKYLSKLEGKIKDNLNKKKLSYLSGLHLLIAELKEV 1160
D 1104 YGKIGEENAOIKDELALARKSKKQOHDCANFVDDLKKSQALDQLTNEKELLISLEO- 1162
QY 1161 IKNNKTYGSPENNTPDNNALLESYKFFLPCTDAVAVVYSSGSDTLEQSQPKRPASTHV 1220
D 1163 --SNSNNEALVEERSDLANRLSDMKKSLSDSNVIVIR--SDLV----- 1203
QY 1221 GAESEITTSQNDDEVNDV-----IIVIFGESEEDVDLGOVYTG----- 1262
D 1204 -----RVNDELDTLKKDKDSISTOYSEVCQDDHDLDSLKGCESEFNKVAASLR 1252
QY 1263 EAVTPSVIDNILSKITENEYEVLYLPLAGVYRSLLKQLENNVMFEN-VNWKDI-LNSRFN 1320
D 1253 ELCTKSEIDVAVSETLDNPFVFNAGNPSLSRLVYLSLENTLAFNOVNFKKMELDNRLT 1312
QY 1321 KRE-NFKNVL-----ESD--LIPYKDLTSSNVYKDPK-FLNKE----- 1356
D 1313 TTDAEFTYVADLEKLEQHEHDMILQRODLEKA--LKDSKKNFLRKAETENIHISLEE 1369
QY 1357 -----KRDKFLSS-----YWKIKSDID--TDINFADVLYGKYIL----- 1389
D 1370 GKETTKETIAELSSKLEDNQATNKLQMDLHNOELRLKEDVLEKESLIIISLESLSN 1429
QY 1390 -SEKYSKDLDS---IKKYINDKQENKERYLPFLNNIETLYKTVNDKI-----DL 1444
D 1430 QROKESLSLDAKNLEHMLDITSRKNS--LMKIEISNSLDDKSELASVETKCAL 1486
QY 1435 FVHLEA-----KVLNITYEKSNN-EVKIKELNLYKTIQDKLADPKKNNFVGADL 1485
D 1487 QKLHSESLSLMENIKSQOEAKEKIOVEDSTIOELDHEITASKNNVEGKDKNDKSIIRD 1546
QY 1486 STDYNH--NNLL-----TKPLST 1501
D 1547 SENTBQNLNLLAEKSAVKRLST 1569

RESULT 12
USOL_YEAST STANDARD; PRT: 1790 AA.
AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE INTRACELLULAR PROTEIN TRANSPORT PROTEIN USOL.
GN USOL OR INT1 OR YDLO584.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=91185402; Pubmed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, usol, is required for intracellular
RT protein transport in Saccharomyces cerevisiae";
RL J. Cell Biol. 113:245-260(1991).
RN [2]
RP SEQUENCE OF 782-1790 FROM N.A.
RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,

RA Kendrick K.E.;
 RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-8 FROM N.A.
 RA Bai Y., Symington L.S.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI COMPLEX.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE ER AND THE GOLGI COMPLEX.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
 CC -1- SIMILARITY: BELONGS TO THE PIL5(TAP)/USO1/YBL047C FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X54378; CA38253.1; -
 DR EMBL: L03188; AAB00143.1; -
 DR EMBL: U53668; AAB6659.1; -
 DR PIR: A38455; A38455.
 DR HSSP: P80220; IDIP.
 DR SGD: S0002216; USO1.
 KW Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil; Calcium-binding.
 FT DOMAIN 1 724 GLOBULAR HEAD.
 FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
 FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
 FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
 FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
 FT CA_BIND 475 486 POTENTIAL.
 FT CONFLICT 847 876 G -> E (IN REF. 2).
 FT CONFLICT 924 924 E -> K (IN REF. 2).
 FT CONFLICT 1253 1253 V -> I (IN REF. 2).
 FT CONFLICT 1319 1319 I -> V (IN REF. 2).
 FT CONFLICT 1461 1581 N -> S (IN REF. 2).
 FT CONFLICT 1581 1581 G -> S (IN REF. 2).
 FT CONFLICT 1600 1600 I -> V (IN REF. 2).
 FT CONFLICT 1661 1661 R -> S (IN REF. 2).
 FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).
 SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

Query Match 5.8%; Score 486; DB 1; Length 1790;
 Best Local Similarity 20.6%; Pred. No. 1.5e-08;
 Matches 314; Conservative 277; Mismatches 562; Indels 368; Gaps 66;

QY 98 GSAVSGSGNSRTPNSDSSDAKSYADIKHVRNLYLTIKELKYPQLEDTNHHMTL 157
 DB 456 GDNAKENGSGNSKESKESDSDKDTGKDTGEYSGFKALFEVL-LNDAEELNLPKLF 514
 QY 158 CDNHGKRYLDS--YEINEL-----LYKNFEDLLRALN--DV--- 195
 DB 515 TTDLFFPFQDDHKYSELEIREITRNVTGNDLEDEEPLKATQITSELTLTSLTAADIRIP 574
 QY 196 -----CANDCQ-----IPNLKIRANELDV--LKKLVFGYRKPLD 229
 DB 575 ISYTLFLIYMFQGFKATNDLSKSVYKLSLFSYQIQQDDVYIKCLVYMLLGAVERF 634
 QY 230 NIKDNVGMEDYIKRKNKT--ENINELIEESK-----TIKKKNA--TKEEK--KR 277
 DB 635 S-KSPSPRKYEFETIKTLGKNVYASRIKQFKKDSYFSDVMEDSLTPPELDETGLPK 693
 QY 278 LYQAOYOLSTYKNOLEKHNLSVLEKRIQTLKKNENIKELDLINELKKNPPANSQTP 337
 DB 694 YVFSTYFIQLFNENI-----YRIPTALSHDPEEPIKSIKIS----- 728

QY 338 NTLIDKNNKIEHEKEIKETIAKIFNIDSLFTDPLDELYLREK-----NKNIDISAKV 392
 DB 729 -----FEVEKLRQRCRK-LKGETISLOTETESTHEULTKLLALNHEKHELODEKY 778
 QY 393 ETKESTEPNEYPCVYTPPLSYNDINNALINELNSFGDLINPDYKREPSKNITYDNKKF 452
 DB 779 QILNSSHSLKENSILETELKKNVDSIDETQRLDRETKD--KENQTALL--EKYST 833
 QY 453 INEKEKIKIEKKIE---SDKSYED-----RSKSLNDITKE---YKLNLEIYDSKFN 502
 DB 834 IHKQDSIKTLEKLETTLSQKKAEDGINKMGKDLPALESKMAVEBNCNKLQKEDKS 893
 QY 503 NIDLTNEKMGKRYKSYVEKLTHTHTFASYENSKHNLEKLTAKLKYMEDYSLNIVYER 562
 DB 894 NVNNOKEKSLKEDLAKITETIKAIN--ENIEEKKIQCNNSKREH-----ISK 941
 QY 563 ELKYRK-----NLISKIENEIETLVENIKRDEQLEFKKITDKPKPDKILEVSDIV 615
 DB 942 ELVEKRSFQSHDNLVAKLTETKLSLANNY-KDMQAEVESLIKAVESKNESSIQLSN-- 998
 QY 616 KVOYQVILAMKIDELKKTQILKKNVELKHNHPNSKQENKQEPYLLVYKKEIDK 675
 DB 999 -----LQNKIDSMQOE-----KENFQIERG-----SIEKNIEQLK 1028
 QY 676 VEMPKVESLINEEKNKIKTEGOSDNSEPTGEITGOATTKPGQAGSALGDSVQAOQ 735
 DB 1029 -----KTIISDLQFQTEIIS--KSSSKDEYTSQI--SLKKEKETATTANDENVKNISE 1079
 QY 736 EOKAQPPVPVPEAKAQVETPPAPVNNKTEVNSKLDYLEKLYEPLNTSYICHKYLVS 795
 DB 1080 LTKTRE-----ELPAEL--AAVKNLKNLETKLETSEKALKEVE----- 1117
 QY 796 HSTNEKTIKQYKTKTEESKLSGCDPLDLNFNTN-----NIPWVSMWD----- 841
 DB 1118 -----NEEHLKEEKTQLEKEAETIQOULNSLANLESKEHEHDLAOLKITEOIANKER 1173
 QY 842 SLNLSOLFMEIYEKEMVNCNLYKLNKDNKIKNLEBAKVKSTSVKTLSSSSMOPLSLTP 901
 DB 1174 QYNELISQNLDELITSTQOENESIK-KKNDELLEGVKNMKSYSSEOSNKKSEIDLALN- 1230
 QY 902 QDKPEVSANDTSTSTNL-----NNSLKT-----PENILSIGKKNKI 938
 DB 1231 QIK-ELKKNKTNEASLIESIKSYSESETVYKIKELQDECNKEFEVSELEDLKASDEKNS 1289
 QY 939 -YOELIOGKSENEYEKILKDSOF-----FYNESFTNPKSADDDINSN--DESKRK 988
 DB 1290 KYLEL--QKSEKIKIEEL--DAKTTELKIQLEKITNLSKAKESSELSRLKKTSSSEK 1345
 QY 989 KLEEDINKLKKTLQISFDLYNKKYKLELRLFDKRRYGVKMQIKLTLLEQLESKXNS 1048
 DB 1346 NAEQOLEKIKNEIOLK-----NQAFEKER-----KLWBGSTTIOEYSEKINT 1389
 QY 1049 LNNPKHVLQNFVFPNKKKEAEIETENTLENTKI---LKHKKGLVKKYNGESSPLK- 1103
 DB 1390 LEDELIRLQNL-----ENELKAKEIOTNRESELEKVSLSNDELLEEKQNTIKSLQDEILSYKD 1445
 QY 1104 --TSESION--BDNVAASLENFVLSKLBGKLDNINLEKKKLSYLSGGLHHLIAELKE 1159
 DB 1446 KITRNDEKLISERQNRDLSLEKQULAAOESKAKYEBGLKLEESSEKKAELERSKE 1505
 QY 1160 VIKKNKNTGNSPSENNDVNNALESYKRFLEPTGDTAVTVVSESGSDTLLEOSQPKPASTH 1219
 DB 1506 MMKRLKSTIES--NEHELKSSMETIK-----SDEKLQES--KKSMEED 1545
 QY 1220 VGAESNTITTSQNVDEVDVYIYPIGESEEDYDGLQVVTGAVPVSYIDNLSIKEN 1279
 DB 1546 I-----KNLOHKSD--LISRIINESKDIIEELKSKRIEAKGSELETVYKQELN 1593
 QY 1280 EYEVLYL-----KPLAGVYRSLK-KOLENNVMFFNVVKDILNSRFKRNENFKNVL 1329
 DB 1594 AQEKIRINAENETVLKSLKLEDIERELDKQAE--IKSNOEEKLLYSRLKELE----- 1644

```

QY 1330 ESDLIPIKDLTSSVYVVDPPYKFLNKKRDLKSLSSYNYINDSIDTIDNFA-----NDVLGY 1385
DQ 1645 -----QELDSFGQAKQK-----SEEBRAEVRKFFQVEKQSDLEKAMLLFTKYNDLVN- 1691
QY 1386 YKILSEKYSKSLDSIKKTYINDKQGENEYKLPFLNNIE---TLTYTVN-----DKIDLFV 1436
DQ 1692 ---KEQAMKREDPVTWKRTTSSQROEIEKLAKELDNLKNKSKLKEANEDESEIDMLMLV 1748
QY 1437 IHLKAVLNTYKESNVEVKI 1457
DQ 1749 TDDEKNKATRSKLKDLGVEI 1769

RESULT 13
RBP1_PLAVB
ID RBP1_PLAVB STANDARD; PRT; 2869 AA.
AC 000796;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
GN RBP1.
OS Plasmodium vivax (strain Belen).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP MEDLINE=92315338; PubMed=1617731;
RX Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
RT merozoites."
RL Cell 69:1213-1226(1992).
CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -1- SUBUNIT: HOMODIMER (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC
DR EMBL: M88097; AAA29743.1; -
DR HSSP: P36956; IAM9.
KW Malaria; Receptor; Signal; Transmembrane.
FT SIGNAL 1 17
FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
FT DOMAIN 18 2807 EXTRACELLULAR.
FT TRANSMEM 2808 2826 POTENTIAL.
FT DOMAIN 2827 2869 CYTOPLASMIC.
FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
SQ SEQUENCE 2869 AA; 330213 MW; B9DBFA42205EBCFF CRC64;

```

```

Query Match 5.7%; Score 477; DB 1; Length 2869;
Best local similarity 20.9%; Pred. No. 4,7e-08;
Matches 372; Conservative 300; Mismatches 617; Indels 490; Gaps 87;

```

```

QY 254 --ELIEESKTTIDKNNKATKEEEKKILYQAOYDLSYKOLEEAAHNLISYLEKRIIDLTK 311
DQ 1115 NKQVISEKVTILFKNNSTV-----YEAHSHINTVAHGI-TSNK 1133
QY 312 NE---NIKELDKIN-----EKNPPANGNTPTLLDKNK-IIEHEKEIKEI 357
DQ 1154 NEILSKYAEVEDKILNLVQNDYKKNKPNENEKQLEAIRGSMKSLKEVINHVSQMGLE 1213
QY 358 --AKTIKFNIDSLFTDPLELEYIREKKN----IDISAKY-----ETKESTEPNEIPNG 406
DQ 1214 STANTLAKSNMAG-----KENHDELELNKTGQMRDIEYKELKIAEELKEGTV- 1261
QY 407 VTYPLSYNDINNALNELNSFGDILNPDTYKPEKKNITDNERKKFNIEKIKIEKK 466
DQ 1262 -----NELKDANEKKNVEP-----EPERNI-----IGHVLERITVEKDK 1296
QY 467 IESDKSYEDRSKSLNDITKYEKELNLIDYDKFNNNIDLTFPKMGRYSYK-VEKLT 525
DQ 1297 A--GKAYEE---MNSLKTIEKILQETSDDS-QNELVTTSTIKHLENMAGYEDVIKRN 1348
QY 526 HHNFPASYENSKHNLEKTLKALTYMEDYSLR-----NIYVELEKTYKNLIS-KIENE 577
DQ 1349 EEDSIQIREKAK-SLETLDKMKKILVQOYNNMIOGASIOGAGISKELMELKGVIELISTN 1407
QY 578 IETLYENIKKDE-----EOLFEKKITK--DENKPEKILLEVSDIVKQVOKVLLMKNID 629
DQ 1408 YSILLEYKKNSSSESVRESOLANGEPFKABEGEKNASARLEAEKLEKQIVKDLIDYSDID 1467
QY 630 E-LKKTQULIKNV-ELKINI-----HVPNSYKQENKQEPYLL----- 664
DQ 1468 DKVKKIBEKIKKELMKKSALTPEWEESEKFKQMCSSHENA--KGGKKIILENNNGDG 1525
QY 665 -----IVLKEIDKLVE-----MKVEESLINE----- 687
DQ 1526 KANITDSQMEEVGVYVSKAEHAFHTVEAOVDKTRAFCEIYAVYTKMDNLFNESLMEKVK 1585
QY 688 ---EKKNKITTEGSDNSRPSRTEGTGATTKPGQOGASALEGSGVQAQAOEQAOQAPV 744
DQ 1586 VKCEKKNDEAKYSKSLKP-YDGR- --KARVSENERKISEL-----KEKAVKRESSQL 1637
QY 745 PVVPEAKAQVTPPAVPANNKTEVNSKIDYLEKLYEFNTSYICHYLVASHSTNEXIL 804
DQ 1638 NDVSTKSLQIDNCRQQLDSVLNIGRVK--QNLQYFDSADKMKSVLPISELGAESL 1695
QY 805 KQYKITYEESKSLSCDPLDLFNIONNIPIWYSMPFDSLNNLSQLEMEIYEKENVCLY 864
DQ 1696 DKVKAKESEYK-----MLETVQNM-----SRINVEEGLNDIDDKITDIEN--DLL 1741
QY 865 KLK---DNDKIKNLLEAKKAVTSVKTILSSSQPLSITPDCKREVSANDTSHSTNLNN 921
DQ 1742 KMKROYEEGLLOKIKENADKRSNFEVLGS-----EINLLDPSTISIFKL 1787
QY 922 SLKTFENILSLGKKKNI-YOELIGO-KSENFYEKILKSDTFYNESFTNFKSKADDIN 979
DQ 1788 KLKEYDMGDL-KNYGVKMEIHGEFTKSYNLITHLSNA-TDVSYVF-----EKAQSLR 1840
QY 980 SL---NDESKRKLLEEDI--NKLK--TQLSFDLYNRYKLEKLELDEKKKTVGYKXMQ 1031
DQ 1841 ELAKKEEHLRRREEAFLFLNDIKKVESLKLKEMMKKVAEYEGMKRDRHTVSQVLQVD 1900
QY 1032 IKKTLLEKQLESKLNLSNNKHYLQNFVYFPNKKKAETAEI-----TEMTLENTYILKHY 1088
DQ 1901 MKTIV---DELKT-LNDISECSSVLNNVSVYKVKESKHADYRRDANSYESAVT--- 1952
QY 1089 KGLVYNGESSPLKTSEESITQEDNVAS---LENFVYLSKLEGLKLDNLLNKKKLSY 1145
DQ 1953 --LANYPLSDAKISSGMEFPAEKSNFKTDLEIFTSVIS-----NSNELLKIEIQ 2002
QY 1146 LSSGILHLIAELKVINKNKNTGNSPSENNTDVNNALSESRYKF--LPEGTDVATVYSES 1202
DQ 2003 DSNVDIQKERESEQLANDA-----TDIYNVVILKLNKNEFNEKLEAKKKEEVYSEK 2051

```

```

QY 1203 GSDPL-----EOSQPKPASTHVGAESENTITTSQNVDEVD 1238
D 2052 VREALKRLSQVEGIRCHENFRLDNTNEELNKKMTIIRDKKSEESOLQEMENMN 2111
QY 1239 DVT-----IYIFGESEDDYDL--GOVVTGEAVTSPVIDNIIISKINEEVLV- 1265
D 2112 TYSNSTOLEGIVVASAESKEDIKLEKRSNEMRNISEKISTIDSKVEMNSTIDELVKL 2171
QY 1286 -----LKLPLAGYVRSK-----KOLENNVMPFNANVAVNLSFRKREKFNVL 1329
D 2172 GKNQCAHMISISTTANMKTISKLLIMKEKENTKCVY- IKDNSSSTGQVETLKGFY 2230
QY 1330 ESDLPYKDLTSSNVYK--DPYKFLNKKRDKFLSSYNTKD-----SIDTDINF- 1378
D 2231 GSKL-----TFSSASEIYQMACTYS-VNPAKHEK--ESLNARLDKKELYLPHQNSDISIV 2283
QY 1379 ---ANDVLGYKILISEKKSLSIDSKYIND-----KOGENEKYL 1416
D 2284 EGVQNMALATDKLNEE-KREMEDELYRNISETKLKOMEHSTVDKFKPMELHGMNETNNK 2342
QY 1417 FLNNIEFLYKTVNDKIDLFVHLEKAVLNTYER-----SNVEYKIKEL----- 1460
D 2343 SLKEKKKLVNDHMSMEKEMKNGKLTYPESYQNNINITYIEAEVKTLEEDRDYG 2402
QY 1461 -NY-----LKTIDKLADEFKNNNFVGIADLSTDYNG----- 1491
D 2403 DNYGIVEBKQFSLIDRTNALMDIEIFKKNY-NLMEVNTETIRVANDYTEKITNK 2461
QY 1492 -----NNLTKFLSTGAVFEKLAR---TVISNLDDNQLQCLN 1526
D 2462 LVQAKTEYQLENIKONDMLQNIPLKVSITIEFENVKRKKESILDYX--QERLLK 2519
QY 1527 ISQH-QCVKQCPONGSCFRLHDERE--CKCLNTYKOE 1562
D 2520 IGEHLDEKRVNVTETLSSY-EIDQKMEMSKNLEKSKS 2557

RESULT 14
MLP1_YEAST
ID MLP1_YEAST STANDARD: PRT; 1875 AA.
AC 00245:
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MYOSIN-LIKE PROTEIN MLP1.
GN MLP1 OR YKR095W OR YKR415.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C.
RX MEDLINE=93247549; Pubmed=8483450.
RA Keelling R., Nguyen T., Chen E.Y., Botstein D.;
RA "A new yeast gene with a myosin-like heptad repeat structure.";
RT Mol. Gen. Genet. 237:359-369(1993).
RL [2]
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94205265; Pubmed=8154186;
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RA "The complete sequence of a 15,820 bp segment of Saccharomyces
RA cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three
RT new open reading frames.";
RT yeast 9.1349-1354(1993).
RL [2]
RN [2]
RP FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
REPAIR.
CC - SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC - CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPL1".
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L01992; AAA34783.1; -
DR EMBL; X73541; CAA51948.1; -
DR EMBL; Z28320; CAA82174.1; -
DR PIR; S38173; S38173.
DR SGD; S0001803; MLP1.
KW Myosin; Heptad repeat pattern; Coiled coil; DNA repair.
FT DOMAIN 69 487 COILED COIL (POTENTIAL).
FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
FT CONFLICT 301 301 R->A (IN REF. 1).
SQ SEQUENCE 1875 AA; 218455 MW; 683A0D34C906867 CRC64;

```

```

Query Match 5.6%; Score 474.5; DB 1; Length 1875;
Best Local Similarity 20.7%; Pred. No. 3.5e-08;
Matches 346; Conservative 289; Mismatches 545; Indels 489; Gaps 78;

```

```

QY 111 TNPDSNDS-DAKSYADL-----KRVNRYLT-IKELYPQLFDLTNNML 155
D 368 TDEADEDENLSAKSSDFIFLKQLIKERTKEHLQOETFIVELE-----HKV 418
QY 156 TLCDNNGFKLIDGYE-EINELLYKLNFFYDILRAKLVQVCADY---CQIFNLKIR 210
D 419 PT---INSFKERTMLELENLNAALLLEHTSNERNKNAVYKELNANQKLECE----- 467
QY 211 ANELDLVAKLTVGYRKPDLNI-----KONVGMEDYIKKNTIENINELIESKRTI 263
D 468 -NDLOTLLKQDLDCRQOYVLLITNSVNSDKPL-----RKKEIQIQIQIMEDDSTI 520
QY 264 DKNNAATKEEKKKLYQAQDLSTYNKOLEBANHLISVLEKRRIDTLK-----KN 312
D 521 -----TESDSQK-----VTERLYEFKNIIQLOEKNAELLVNRLADLESKE 564
QY 313 ENIKELDKI-NEIKNPPRANSNTPTLLDKNKKIEHEKEIKETIKTFINISLFTD 371
D 565 KKSQSLQKLESETVNEAKKAITLKSEKMDLESRIEQLQELBEL-KTSPVNDASYS- 622
QY 372 PLELEYLRKKNNIDISAKVETKESTEPNEYPGVYTPLSYNDJNNALNLSFGDLIN 431
D 623 -----NVTIKQLETFRKLES-----QVQDLQIRISQI----- 650
QY 432 PFDYTKPSKNIYDNERKKFINDIEKIKETKKIESDKKSYDRSKSLNDITKEYEKL 491
D 651 ---TRESTENMSLLNKEIODLYPSKDISIKLQEKSSRIAEERFKLL----- 696
QY 492 LNEYDSKFNNDLITNEK-NMGKRSY-----KYEKLTHNTPFASYENKHNLEKL- 543
D 697 -----SNTDLITRAENDQLRRREDYLONTILKQDSKT-HETLNEYVSCSKSLIVE 746
QY 544 TKALKMEDYSLRNIVVEKELKYRNLSKIENEITLV--ENIKRDEQLFPEKTKTKD 600
D 747 TELNLKEEQKLR-VHEKKNLQKELNKLSPKSDSLRIMVQLQTLQEREDLLETKSC 805
QY 601 ENKPDEKILEVSDIVKQVOK-----VLANKIDELK-----TQILK 639
D 806 OKKIDLEDELALSELKETSQKDHKIKOLEEDNNSNIEMYOQKRIALKRDYESVITSYDSK 865
QY 640 NVELKHNHIVNSYKQENKQEPYLLV-----LKKELDKLVFAPKYESLINE 687
D 866 QTDIEKLQYKVSLEKLEEDKIRLHYNVWDETINDSLKELKESKINLTDAYSQIKE 925
QY 688 EKKNIKTEGSDNSEPSTEGEITGQATTKPGQAGSALLEGSDVOAQOEOQKQAPVPVP 747
D 926 YKDLTETTSQ-----LQQTNSKLD-ESFEDFLNQIKNL----- 958
QY 748 VPEAKQVPPPPAPVNNKTENVSKLDYL-EKLYEFLNTSYTCHRYLIVSHSTNKEKL- 804

```

```

Db 959 -----TDEKTSLEDKISLKEQMFNLNNELDLQKKMEKAKFKKRISIL 1004
Oy 805 ----KOYKITEE-ESKLS-----SCDPLDLFN-IONNIPMYGMSFDSLNLSISQLEFMEI 854
Db 1005 ONNKEVANKSEYESKLSIQNDLDOOTIYANRANNYEOELOKHADSVKSTISELREOL 1064
Oy 855 YEKEMVONLYKLNKNDKIKNLLEAKKYSTSVK-----TLSSSMQPLS-----LTPQ 902
Db 1065 HTYGQVKTNL-SRDQLENLKENEKSMWSQKESLLEQDLNSRIEDLSQKKLTYDQ 1123
Oy 903 DKPEVSANDTSHSTN--LNSSLKLEFNILSLCKNNIY--QELIGKSSSENFYEKT- 955
Db 1124 IQITLADKVENNSTNGPLNNIL-----ITLREREDILDTKYTAERDAKMLROKISTL 1177
Oy 956 ----LKSDPFYNSFT-----NFVKSADINDSL-----NDESKR 987
Db 1178 MDVELQDAPKLDNSRVEKENHSSIIQOHDIMEKLNOLNLRNSNTTLRNELENNNKK 1237
Oy 988 KLEBEDINKLKTLL-----QLSPDLYNK-YKLTLELFDKKKTKYKMOIKITLL 1038
Db 1238 KELOSELDKLONVAPIESELTLAKYSMOEKEQLK-----AAEYHAKKKRSODILEK 1292
Oy 1039 KEQLES-----KLNS-LNNPKHVLQN-----FSVFENK-KKEAE-----TAE 1073
Db 1293 HEQLSSDYERLESEIELEKLEENKEROGAEADEKFNRLRQOERLTKTSKLSQDSLTE 1352
Oy 1074 TENTLEMTKLLKTKYKGLVYKNGESSPLKLSPESTIOTEDNTASLENFKYK---SKLEG 1130
Db 1353 QVNSLRKAKVLEN-----SLSEANARIEE-----LONAKVAQGNQMLEA 1392
Oy 1131 --KLKDN-----LNLEKKLSYLS--SGLHLLIAELKEVIR-----NKNTGNSPS 1172
Db 1393 IRKIQDEAKASRLQAKLEESTISTYSTINGLNEEITTLKEELEKORQIOQLOANSAN 1452
Oy 1173 ENNTDVNNALESYKPLPEGTIVATVYVSESGDYLE--OSQPKPASTHGAESNTITT 1229
Db 1453 EQN-DLSNIVESMKKSFEE--DKIKFIKEKTOEVNEKILEAQERLNQPSNINMEIKKKW 1509
Oy 1230 SQNVDEVDVYIPIGESESEYDQGVYTGAVTPPSVINDILSKITENEYVLYAKPL 1289
Db 1510 ESEHEQVSOKI-----REAEELAKKRIRLPTTEKIN-KIIEKKKEELEKEFEKKEVERE 1563
Oy 1290 AGYRS-----LKQLENNVMTFNVMVKDILNSRFNKRNFKNVLSDDLTPYDLTSSN 1343
Db 1564 KSMQOSEIDVYLRKQLEAKVQ--EKQKLENEYKK-----LOBEL--KDVPHSS 1610
Oy 1344 YVVKDPYKFLNKKERKDFLSSYNIKDSITDIDINPANDVLGYKILSEKYSKSDLSIKKY 1403
Db 1611 HISDD-----ERDK-----LRAEISR-----LREFEANNELQAIKKK 1642
Oy 1404 INDKOGENEKLPLPLNNIEITLYKTVNDKIDLFVHLEAKVLTNYEKSNEVAKIKELNYL 1463
Db 1643 SFD-EGKO-----QAMMKTT-----LLEKK-----L 1662
Oy 1464 KTIDDKLADFKNNNFYGIADLSTDYNNHNLTKFLSTGAVFENLAKTVLSNLDGNSLG 1523
Db 1663 AKMSQSLSEKQSAE-----SPPKSVNVQNPBLGLPKRIEENSNSPPPLLSG----- 1711
Oy 1524 MLNLSHQCVKQCPQNSGCFRLDERECCCLNLYKQEGDKVENDNP 1572
Db 1712 -----EKLTKLNSKSSSGGFNPFTSPSPKHLQNDNDKRSLANKNIDP 1754

```

```

RESULT 15
RBP2_PLAVB
ID RBP2_PLAVB STANDARD; PRT; 1251 AA.
AC 000799;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE RETICULOCTE BINDING PROTEIN 2 (FRAGMENT).
GN RBP2.
OS Plasmodium vivax (strain Belém).

```

```

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9231538; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
RL Cell 69:1213-1226(1992).
CC -1- FUNCTION: INVOLVED IN RETICULOCTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCTE CELLS.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND (PROBABLE).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M88098; AAA29744.1; -
KW Malaria; Receptor; Membrane.
FT NON_TER 1
FT NON_TER 1251
SQ SEQUENCE 1251 AA; 143741 MW; 54BA51C740AC572 CRC64;

```

Query Match 5.4%; Score 457.5; DB 1; Length 1251;
Best Local Similarity 21.1%; Pred. No. 7,7e-08;
Matches 306; Conservative 226; Mismatches 458; Indels 455; Gaps 71;

```

Oy 272 EEEKKLYQAYDYSYKQLEAHNLISYLEKRIDLKKNENIKELDKITENIKNPPRA 331
Db 1 EFDEKVEDTSFDEK--KKSIEKAY-----EKKMGNTLKE-----LEKMDKEN-- 41
Oy 332 NSGMPNPLDKNKKIEPHEKEIKETIKTFNPD--SLTPDPLELEYLREK--NK 384
Db 42 -----TEKEVEEAQIOYKRIF-----IDHDVNLNDEVEKSKIMEKIELYKK 84
Oy 385 NIDISAKVETKESTEPNYPNGVT-----YPLSYNDINNALNELNSFGDLINPDTYKPS 440
Db 85 EID-----EIQKKT--NEYKQDTSNFPYTTQYNSAQSKAKIGOFINATTKGTSTPS 137
Oy 441 KNITDNERKKFNIKEIKIEKKIKIESDKSYEDRSK--SLNDITKEYEKLIN--EI 495
Db 138 QDI--NE--LESIKEVEHKNQLVYQESNSMEEMKQILSMKDLL--ILNSET 185
Oy 496 YDSFNNNIDILTFNEKMMGKRYKVEKLTGHNTPASTENSKHNLKTLAKTYMEDSL 555
Db 186 IAKETISNT-----QNALGFEENKTKLNTKIDELLO----- 216
Oy 556 RNIVVEKELKYKKNLISKIENEIEFLIVENIKKDEEQLFEKKITRQDENKPRDEKILEVSDIV 615
Db 217 RVAAIMEBAKHK-----NNIDIALEDAQID-----TEVSKIE 249
Oy 616 KVQVQVYLLMKIDELKKTQILKLNVELKHNHVPNSYKQENKQEPYVLYLVAKKEIDIK 675
Db 250 QINRE--INNKKDEIKSYLSEIKYEKDKCTTEISNSRGKDK-----IEPLE 294
Oy 676 VFMKVES-----LINEKKNIKTEGQ-----SDNSPSTEGEITGOATTKPQOQAGSA 724
Db 295 KFKPNEESNSKVNINENINENIRNSQYLKIDIEAERKQASTKVELFHKHEF--TISNIF 351
Oy 725 LEGSDVOAQAEQKQAPVVPVPEAKAQVTPPAPVNNKTENVSKLDYLEKLYEFLNT 784
Db 352 KESRILGVEYTSQKKI-----NKAEDIMK--ELER----- 379
Oy 785 SYICHKYLIVSHSTMNEKILKQYKTKREESKLSGCDPLDLFNIONNIPYVY--SMEDSL 843
Db 380 -----HNSEIQTQVKGFOENLKNL-----NEPHANYNAEDEL 411
Oy 844 NNSLSQLEFMEIYKEMVCNLYKLNKDN-DKIKNLLEAKKYSTSVKTLSSSMQPLSLTPQ 902

```


Db 412 NNDKSTNAKVLE-----TNLESVKNHLSITNKGGEIKYSKAKDI-----MOKIKATSE 463
QY 903 DKPEVS---ANDDTSHTNLNNSLAKFENILSLGKN-----NIYOLLIGOKSSEN 950
Db 464 NTAKETLEKVDQDSNVYVNLQITTERNLVTETKRNRLNGIDSTITNEGAL--KESKG 520
QY 951 FYE-----KILKSDTFYNESTFNVKSKADDINSLAND-ESKR 987
Db 521 NYETGFLEKLEELIGKNRKLKVDITKKSINSTVGNFSSLFNNFDLNOYDFKNKINDYENKM 580
QY 988 ---KLEBEDINKLKTLOLSFDL---YNKYK-LKLE-----RPFDKKTVGKYKMOIK 1033
Db 581 GEIYNFEESLKNISENIRNASENTSDNSAKTLRLLEAKOKEKVNILNKEEANKRYLRDVK 640
QY 1034 KLTL-----LKEOLESKLNSL-----NNPKHVLONFVSFFNKKKEAEIAE-- 1073
Db 641 KVESFRFIFNNKESLD-KINEMIKKEOLTVNEGHGNVQOLVENIKELVDENNLSIDLKQA 699
QY 1074 -----TENTLEN-TKILKHVKGLVRYNGESSPLKLTSEESIOTEDNYASLENF 1122
Db 700 TGNKEEIOKITHSLKKNKAKIILGHVDSAKYVGKITPELALTE----- 744
QY 1123 KVLKSLBGKLDKNILNLEKKKILSYSSGLHLIAELKEVYKKNKNTGNSPSENNTDVN-NA 1181
Db 745 -LLGDALKETPAQELKPFESK-----NNVLETEMNSKN-----TNELDVHKNI 785
QY 1182 LESYK---KFLPEGTDVATVVSSESG-----SDTLEQSQPKKPA-ST 1218
Db 786 QDAKVALEILIAHDEIDTKOKDSSKLIEMGNQIYLKAVLINQYKKNKISSIKSKEEAVSV 845
QY 1219 HVG-----AESNTITTSQNVDEVDVYIPIFGESEEDYDLGOVVTGEAVTPSYIDN 1272
Db 846 KIGVNSKKHSELKITS---DKSYDNIAL---EKQTELONLNRSPTQEK-TINTSDS 897
QY 1273 ILSKIENEYEVL--YLRPLAGVYRSIKQOLEN--NWTFFNVNVKDIILNSRPKNRENEKNV 1328
Db 898 KLERIKTDPELSLKNALKTLEGEVNALKASSDNHSHVQSKSEPVNPAL-SEIEKEET--- 952
QY 1329 LESDLIPYKDI/SSSNVYVVKDPYKFLNKEKRPDPLSSVYIKDSI-----DTDI--NEAN 1380
Db 953 -----DIDSLNTALDE---LKKGRICEVSRKLIKDVYKELISDDTELINTIEK 999
QY 1381 DVLCGYRILSEKYR---SDLDISIKYINDKOGENEKYLPE-----LNNIETLYKTV 1428
Db 1000 NVKAYLAYIKKNYEDYQDVYLTNEHFTKQVSNHEPTNFDKSNKSSSEELTKAVTDSKTI 1059
QY 1429 NDKIDLEVILHLEAVLVNTYEKSNEVYKIKEL---NYLKTIODLADFKKNNNFVGIAID 1484
Db 1060 ISKLKGVIIE---VVENTEMNTIESSAKIEALYNELKNKTSLSNEIYOTSNEVKLOE 1114
QY 1485 LSTOYNHNNLTFLSTGWFENLAKTVLSNL-----LDGNSLOMLNISQH- 1530
Db 1115 MKNAD-----KIYDVKIPIFTVDTOKSNITVNOHSINNWKDLKGLKLOELIDADSSF 1168
QY 1531 --QCVRK 1535
Db 1169 TLESIRK 1175

Search completed: October 27, 2001, 15:57:18
Job time: 248 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2001, 15:49:05 ; Search time 36.52 Seconds
(without alignments) 3418.677 Million cell updates/sec

Title: US-09-269-874-3
Perfect score: 8424
Sequence: 1 MKIIFELCSFLFIINTQCV.....SNLGLISFLILMLILYSFI 1639

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR-68:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	8424	100.0	1639	2	S05603	major merozoite su
2	8191.5	97.2	1640	2	A24594	probable major sur
3	7965	94.6	1631	1	SAZOK1	major merozoite su
4	5136.5	61.0	1726	1	SAZOGM	major merozoite su
5	5121.5	60.8	1726	2	A45948	major merozoite su
6	5065	60.1	1701	2	A54498	major merozoite su
7	5053	60.0	1701	2	A26868	major merozoite su
8	3093.5	36.7	1751	2	A45604	major blood-stage
9	2999.5	35.6	1086	2	S16752	major merozoite su
10	2959	35.1	1726	2	A39401	merozoite surface
11	2829.5	33.6	1060	2	S06286	major merozoite su
12	2548.5	30.3	1772	2	A45532	major merozoite su
13	2509	29.8	1785	2	A45346	major merozoite su
14	2146	25.5	651	2	S47282	merozoite surface
15	2101	24.9	400	2	A45545	major merozoite su
16	1187	14.1	233	2	A25814	glycoprotein 185 -
17	1135	13.5	680	2	A28121	major merozoite su
18	1096.5	13.0	281	2	B54498	major merozoite su
19	1021.5	12.1	636	2	A45949	merozoite surface
20	815	9.7	168	2	A25522	major merozoite su
21	710.5	8.4	146	2	S03290	glycoprotein, 190K
22	659.5	7.8	138	2	P00125	major merozoite su
23	606.5	7.2	1939	2	T18372	repeat organellar
24	605	7.2	2401	2	T28676	riophy protein -
25	568.5	6.7	2269	2	T28677	riophy protein -
26	559.5	6.6	1979	2	C71622	hypothetical prote
27	529	6.3	115	2	B25120	major merozoite su
28	521.5	6.2	3394	2	T18501	hypothetical prote
29	507	6.0	2166	2	G70163	hypothetical prote

30	501.5	6.0	1957	2	T38077	hypothetical colle
31	495.5	5.9	284	2	A45518-	merozoite surface
32	487	5.8	1790	2	S67593	transport protein
33	483	5.7	1127	2	T28317	ORP MSV156 hypothe
34	477	5.7	2829	2	A42771	reticulocyte-bind
35	477	5.7	3724	2	T18427	hypothetical prote
36	474.5	5.6	1875	2	S38173	myosin-like protei
37	465	5.5	476	2	B44997	merozoite surface
38	460	5.5	482	2	A44997	merozoite surface
39	457.5	5.4	1252	2	B42771	reticulocyte-bind
40	453.5	5.4	2485	1	H71621	serine/threonine-s
41	448	5.3	1650	2	T18444	hypothetical prote
42	443.5	5.3	2663	1	S28261	centromere protein
43	442.5	5.3	3973	2	B71612	hypothetical prote
44	440	5.2	1805	1	A64224	hypothetical prote
45	438	5.2	1738	2	T14867	interaptin - slime

ALIGNMENTS

RESULT 1
S05603 major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
N:Alternate names: gp195 surface antigen
C:Species: Plasmodium falciparum
C>Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jun-2000
C:Accession: S05603; S04850
R:Myler, P.J.
submitted to the EMBL Data Library, April 1989
A:Reference number: S05603
A:Accession: S05603
A:Molecule type: mRNA
A:Residues: 1-1639 <MYL>
A:Cross-references: EMBL:X15063; NID:99896; PIDN:CAA33163.1; PID:99897
R:Myler, P.J.
Nucleic Acids Res. 17, 5401, 1989
A:Title: Nucleotide and deduced amino acid sequence of the gp195 (M5A-1) gene from P1
A:Reference number: S04850; MUID:89345116
A:Accession: S04850
A:Molecule type: mRNA
A:Residues: 1504-1639 <MYL2>
A:Cross-references: EMBL:X15063
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1639/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 100.0%; Score 8424; DB 2; Length 1639;
Best Local Similarity 100.0%; Pred. No. 6.2e-254;
Matches 1639; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MKIIFELCSFLFIINTQCVTHESYQELVKRLAEADVLNGYSLFQEKWLNIGTSGT	60
DB	1	MKIIFELCSFLFIINTQCVTHESYQELVKRLAEADVLNGYSLFQEKWLNIGTSGT	60
QY	61	AVTSTPGSKGSVAGSGGSVAGSGSVAGSGSVAGSGSVAGSGSGNSRTPNSDSS	120
DB	61	AVTSTPGSKGSVAGSGGSVAGSGSVAGSGSVAGSGSVAGSGSGNSRTPNSDSS	120
QY	121	DAKSYADLKHRYRNVLTKTELKYPQLFDLNNHMLTLCDNHIGFRYLLDGYEEINELLYK	180
DB	121	DAKSYADLKHRYRNVLTKTELKYPQLFDLNNHMLTLCDNHIGFRYLLDGYEEINELLYK	180
QY	181	LNFRYDILLRAKINDVCANDYCOIPNLEKIRANEDLVLLKLVFGYRKPDPDNKDNVGMED	240
DB	181	LNFRYDILLRAKINDVCANDYCOIPNLEKIRANEDLVLLKLVFGYRKPDPDNKDNVGMED	240
QY	241	YIKKKKTIENINELIEESKKTIDKNKNATKEEKKKLYOAOYDLSTYNKOLEEAMHLS	300
DB	241	YIKKKKTIENINELIEESKKTIDKNKNATKEEKKKLYOAOYDLSTYNKOLEEAMHLS	300

Qy 301 VLEKRIDTLKKNENIKELLDKINEIKNPPANSNGTPTLLDKKKIEHEKEIKETAKT 360
Db 301 VLEKRIDTLKKNENIKELLDKINEIKNPPANSNGTPTLLDKKKIEHEKEIKETAKT 360
Qy 361 IKFNIDS.LFTDPLEEYLLREKKNKIDISAKVETKESTEPNEYNGVYPLSYNDINNAL 420
Db 361 IKFNIDS.LFTDPLEEYLLREKKNKIDISAKVETKESTEPNEYNGVYPLSYNDINNAL 420
Qy 421 NELNSFGDLINPEDYTKPSKNITTDNERKKFINEIKKIKIEKKIESDKKSYEDRSKS 480
Db 421 NELNSFGDLINPEDYTKPSKNITTDNERKKFINEIKKIKIEKKIESDKKSYEDRSKS 480
Qy 481 LNDITKEYEKLINLEIYDSKFNNNIDLTNEFKMKGRYSYKVEKLTHTNFASYSKSNL 540
Db 481 LNDITKEYEKLINLEIYDSKFNNNIDLTNEFKMKGRYSYKVEKLTHTNFASYSKSNL 540
Qy 541 EKLTKALKYMEDYSLRNIVVEKELKYKNLSKIENEIETLVENIKKDEQLFEKKIKTD 600
Db 541 EKLTKALKYMEDYSLRNIVVEKELKYKNLSKIENEIETLVENIKKDEQLFEKKIKTD 600
Qy 601 ENKPDEKILEVSDIYKVOVKVLLMKNIDELKTKQULIKNVELKHNHVPNSYKOEKOE 660
Db 601 ENKPDEKILEVSDIYKVOVKVLLMKNIDELKTKQULIKNVELKHNHVPNSYKOEKOE 660
Qy 661 PYYLLIVLKEIDKLVEMPKVESLINEEKNKITEGOSDNSEPTCEBITGOATTKPGQ 720
Db 661 PYYLLIVLKEIDKLVEMPKVESLINEEKNKITEGOSDNSEPTCEBITGOATTKPGQ 720
Qy 721 AGSALGEDSVQAQOEQAOPPVYPPVPEAKAOVPTPAPVNNKTEVNSKIDYLEKLYE 780
Db 721 AGSALGEDSVQAQOEQAOPPVYPPVPEAKAOVPTPAPVNNKTEVNSKIDYLEKLYE 780
Qy 781 FLNLSYIOCHKIYLVSHTSMNEKILKOYKITKEESKISSCPDLLEFNIQNNIPMYSMF 840
Db 781 FLNLSYIOCHKIYLVSHTSMNEKILKOYKITKEESKISSCPDLLEFNIQNNIPMYSMF 840
Qy 841 DSLNNSLSQLEFMEIYKEMVCNLYKLDNDKIKNLEBAKRVSTSVKTLSSSSMQLSLT 900
Db 841 DSLNNSLSQLEFMEIYKEMVCNLYKLDNDKIKNLEBAKRVSTSVKTLSSSSMQLSLT 900
Qy 901 PODPREVANDTSHSTNLNNSIKLFENILSLGKNKNTYOELIGQKSENEYEKLKXSD 960
Db 901 PODPREVANDTSHSTNLNNSIKLFENILSLGKNKNTYOELIGQKSENEYEKLKXSD 960
Qy 961 TFYNESEFNFYKSKADINSINDESKRKLEEDINKLKTQLSFDLYNKKLKERLFD 1020
Db 961 TFYNESEFNFYKSKADINSINDESKRKLEEDINKLKTQLSFDLYNKKLKERLFD 1020
Qy 1021 KKTIVGKYMQIKKLTLLKEQLESKLSLNNPKHVLQNFVFNKKKEBAEIAETENTLEN 1080
Db 1021 KKTIVGKYMQIKKLTLLKEQLESKLSLNNPKHVLQNFVFNKKKEBAEIAETENTLEN 1080
Qy 1081 TKILLLHKGLVKKYNGSSPLKLTSEESIOTEDNYASLEMFKYLSKLEGLKDNLEK 1140
Db 1081 TKILLLHKGLVKKYNGSSPLKLTSEESIOTEDNYASLEMFKYLSKLEGLKDNLEK 1140
Qy 1141 KKLSTYSSGLHLLAELKEVYKKNKYTNSPENNVDNNALESKPKLPCTDVAIYVS 1200
Db 1141 KKLSTYSSGLHLLAELKEVYKKNKYTNSPENNVDNNALESKPKLPCTDVAIYVS 1200
Qy 1201 ESGSDTLEQSQPKRPASTHVAESNTITTSQNVDEVDVILVPLFGESEBDYDGLGVV 1260
Db 1201 ESGSDTLEQSQPKRPASTHVAESNTITTSQNVDEVDVILVPLFGESEBDYDGLGVV 1260
Qy 1261 TGEAVTPSVIDNITLSKIENEVEVLYLKLAVYNSLAKOLENNWTFVANNKDLINSRPN 1320
Db 1261 TGEAVTPSVIDNITLSKIENEVEVLYLKLAVYNSLAKOLENNWTFVANNKDLINSRPN 1320
Qy 1321 KREPNKVNLESDDLPIYKOLTSNNVVDYKPYFLNKEKRDKFLSSVNYIKDSDIDIDINAN 1380
Db 1321 KREPNKVNLESDDLPIYKOLTSNNVVDYKPYFLNKEKRDKFLSSVNYIKDSDIDIDINAN 1380
Qy 1381 DVLGYKILSEKYSDDLSIKKIYNDKOGENEKYLPLFNNIETLYKTVNDKIDL.FVHLE 1440

Db 1381 DVLGYKILSEKYSDDLSIKKIYNDKOGENEKYLPLFNNIETLYKTVNDKIDL.FVHLE 1440
Qy 1441 AKVLNTYKESNVEKIKTELNYLKTIOQKLDKFKKNNNFVGIADLSDYNNNNLTFLS 1500
Db 1441 AKVLNTYKESNVEKIKTELNYLKTIOQKLDKFKKNNNFVGIADLSDYNNNNLTFLS 1500
Qy 1501 TGMVFENIAKTIVLSNLDGNGLMNISQHOVCVKKOCPONSGCFRHLDERECKLLANYK 1560
Db 1501 TGMVFENIAKTIVLSNLDGNGLMNISQHOVCVKKOCPONSGCFRHLDERECKLLANYK 1560
Qy 1561 QEGDKVNEPNPTCNENNGGCDADAKCTEEDSGSNKKITCECTKPSYPLFDGIFCSSS 1620
Db 1561 QEGDKVNEPNPTCNENNGGCDADAKCTEEDSGSNKKITCECTKPSYPLFDGIFCSSS 1620
Qy 1621 NFIQSFLILMLITYSFI 1639
Db 1621 NFIQSFLILMLITYSFI 1639

RESULT 2

A24594
probable major surface antigen (83K, 19K, 42K) precursor - malaria parasite (Plasmodi
C:Species: Plasmodium falciparum
C:Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 09-Jun-2000
C:Accession: A24594
R:Holder, A.A.; Lockyer, M.J.; Odink, K.G.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholl
Nature 317, 270-273, 1985
A:Title: Primary structure of the precursor to the three major surface antigens of Pl
A:Reference number: A24594; MUID:86014355
A:Accession type: DNA
A:Residues: 1-1640 <HOL>
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 97.2%; Score 8191.5; DB 2; length 1640;
Best Local Similarity 97.9%; Pred. No. 9.8e-247;
Matches 1605; Conservative 8; Mismatches 26; Indels 1; Gaps 1;

Qy 1 KMIIFELCSLFFIINTOCVTHESYQELVKLEDAEDAVLTGYSLFQEKKNVLENGTSGT 60
Db 1 KMIIFELCSLFFIINTOCVTHESYQELVKLEDAEDAVLTGYSLFQEKKNVLENGTSGT 60
Qy 61 AVTSTPSKGSVSGSGSVASGGSVSGSVASGGSVSGSGSNRRTPSPDSSDS 120
Db 61 AVTSTPSKGSVSGSGSVASGGSVSGSVASGGSVSGSGSNRRTPSPDSSDS 120
Qy 121 DAKSYADLKHVRNVLTLTIKELKYPQLFDTLNNHMLTLCDNTHGRKYLLIDGEEINELLK 180
Db 121 DAKSYADLKHVRNVLTLTIKELKYPQLFDTLNNHMLTLCDNTHGRKYLLIDGEEINELLK 180
Qy 181 LNFYFDLIRAKLNDVCANDYCOIPNMLKIRANEDLVKLLVGYGRKPLDNIKDNVGMED 240
Db 181 LNFYFDLIRAKLNDVCANDYCOIPNMLKIRANEDLVKLLVGYGRKPLDNIKDNVGMED 240
Qy 241 YIKKNKTKIENINELIESKKTIDKNNKATKEEEKKKIYQAOYDLSYNNKOLEAHNLIS 300
Db 241 YIKKNKTKIENINELIESKKTIDKNNKATKEEEKKKIYQAOYDLSYNNKOLEAHNLIS 300
Qy 301 VLEKRIDTLKKNENIKELLDKINEIKNPPANSNGTPTLLDKKKIEHEKEIKETAKT 360
Db 301 VLEKRIDTLKKNENIKELLDKINEIKNPPANSNGTPTLLDKKKIEHEKEIKETAKT 360
Qy 361 IKFNIDS.LFTDPLEEYLLREKKNKIDISAKVETKESTEPNEYNGVYPLSYNDINNAL 420
Db 361 IKFNIDS.LFTDPLEEYLLREKKNKIDISAKVETKESTEPNEYNGVYPLSYNDINNAL 420
Qy 421 NELNSFGDLINPEDYTKPSKNITTDNERKKFINEIKKIKIEKKIESDKKSYEDRSKS 480
Db 421 NELNSFGDLINPEDYTKPSKNITTDNERKKFINEIKKIKIEKKIESDKKSYEDRSKS 480

QY 481 LNDITREYKELINEIYDSKFNNNIDLTNFEKMGKRSYKVEKLTHTNFASYSKHNL 540
D 481 LNDITREYKELINEIYDSKFNNNIDLTNFEKMGKRSYKVEKLTHTNFASYSKHNL 540
QY 541 EKLTKALKMEDYSLRNIVVEKELKYKKNLSKISKENIETLVENIKKDEDLFEKKITKD 600
D 541 EKLTKALKMEDYSLRNIVVEKELKYKKNLSKISKENIETLVENIKKDEDLFEKKITKD 600
QY 601 ENKDEKILEVSDIVKVOQVLLMNKIDELKTKQLILKNELHNNHVPNSYQENKOE 660
D 601 ENKDEKILEVSDIVKVOQVLLMNKIDELKTKQLILKNELHNNHVPNSYQENKOE 660
QY 661 PYLIVLAKKEIDKLVMPKVESLITNEEKNNIKTEGOSDNSEPECEGITTQATTKPCOO 720
D 661 PYLIVLAKKEIDKLVMPKVESLITNEEKNNIKTEGOSDNSEPECEGITTQATTKPCOO 720
QY 721 AGSALBGDSVOAQAOEQKQAPVPVPVPEAKAQPVPVPVNNKTENVSXLDYLEKLYQ 780
D 721 AGSALBGDSVOAQAOEQKQAPVPVPVPEAKAQPVPVPVNNKTENVSXLDYLEKLYQ 780
QY 781 FLNMSYICHKIYLVSHSTMNKILKQYKITEESKLSGCCPDLPLNTIONNIVYMSMF 840
D 781 FLNMSYICHKIYLVSHSTMNKILKQYKITEESKLSGCCPDLPLNTIONNIVYMSMF 840
QY 841 DSLNNSLSOLFMEIYKEMVCNLKLDNDKIKNLLEBAKKVSTSVKTLSSSMOPLST 900
D 841 DSLNNSLSOLFMEIYKEMVCNLKLDNDKIKNLLEBAKKVSTSVKTLSSSMOPLST 900
QY 901 PODKPEVASANDTSHSTNLNNSLKLFEINILSLGKNKNIOELIIOKSESENYEKLKDS 960
D 901 PODKPEVASANDTSHSTNLNNSLKLFEINILSLGKNKNIOELIIOKSESENYEKLKDS 960
QY 961 TFYNESTNFKYKADDIINSLNDESKRKLLEEDINKLTKTQLSFDLYNKKYKLERLD 1020
D 961 TFYNESTNFKYKADDIINSLNDESKRKLLEEDINKLTKTQLSFDLYNKKYKLERLD 1020
QY 1021 KKKVGVYKMKOIKKLTLLKEOLESKLNSLNPKHYLONFSEFNKKEAEIETENLEN 1080
D 1021 KKKVGVYKMKOIKKLTLLKEOLESKLNSLNPKHYLONFSEFNKKEAEIETENLEN 1080
QY 1081 TKILKHYKGLVKNYNGESSPLKTLSESIOTEDNYASLENFYVLSKLECKLKNLNEK 1140
D 1081 TKILKHYKGLVKNYNGESSPLKTLSESIOTEDNYASLENFYVLSKLECKLKNLNEK 1140
QY 1141 KKLSTYSSGLHLLIAELKELYKKNKNTGNSPEKNTDVNNALESYKFLPBGDTVAAYVS 1200
D 1141 KKLSTYSSGLHLLIAELKELYKKNKNTGNSPEKNTDVNNALESYKFLPBGDTVAAYVS 1200
QY 1201 EGSGLTLEOSOPKPPASTHVGAESENTITTSQNVDEVDVITVIFGSEEDYDLDGOV 1260
D 1201 EGSGLTLEOSOPKPPASTHVGAESENTITTSQNVDEVDVITVIFGSEEDYDLDGOV 1260
QY 1261 TGEAVTTSVIDNITLSKIENEYEVLYLPLAGVYRSKLEKOLENNVMTFNANVKKDLNRSFN 1320
D 1261 TGEAVTTSVIDNITLSKIENEYEVLYLPLAGVYRSKLEKOLENNVMTFNANVKKDLNRSFN 1320
QY 1321 KREFFKAVNLESDLIPYODLTSSNVAVDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFA 1380
D 1321 KREFFKAVNLESDLIPYODLTSSNVAVDPYKFLNKEKRDKFLSSYNYIKDSIDTDINFA 1380
QY 1381 DVLGYYITLSEKYSKSDLSIKKYINDOGENEKYLPLNNIETLYKTVNKKIDLFVHLE 1440
D 1381 DVLGYYITLSEKYSKSDLSIKKYINDOGENEKYLPLNNIETLYKTVNKKIDLFVHLE 1440
QY 1441 AKVLYNTYKESNVEVKIKELNYLKTIDOKLADFPKNNNFVGIADLSTVDYNNHNLTKFELS 1500
D 1441 AKVLYNTYKESNVEVKIKELNYLKTIDOKLADFPKNNNFVGIADLSTVDYNNHNLTKFELS 1500
QY 1501 TGWFFEWLATTIVSLNLDGNL-OGMLNISOHCYKACPONSGCFRULDERECCILANT 1559
D 1501 TGWFFEWLATTIVSLNLDGNL-OGMLNISOHCYKACPONSGCFRULDERECCILANT 1559
QY 1560 KOEDKCVENPNTCENNGGCADADAKTEEDSGSNCKKITCECTKRDPSYPLFGICSS 1619
D 1560 KOEDKCVENPNTCENNGGCADADAKTEEDSGSNCKKITCECTKRDPSYPLFGICSS 1619

Db 1561 KQESKCVENPNTCENNGGCADADAKTEEDSGSNCKKITCECTKRDPSYPLWVICSS 1620
QY 1620 SNFJGISFLILMLILYSFI 1639
D 1621 SNFJGISFLILMLILYSFI 1640
RESULT 3
SAZOKI
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 09-Jun-2000
C:Accession: A25120
R:McKee, M.; Goman, M.; Bone, N.; Hyde, J.E.; Scaife, J.; Certa, U.; Stunnenberg, H.
EMBO J. 4, 3823-3829, 1985
A:Title: Polymorphism of the precursor for the major surface antigens of Plasmodium f
A:Reference number: A91030; MUID:86136024
A:Accession: A25120
A:Molecule type: DNA
A:Residues: 1-1631 <MAC>
C:Comment: The merozoite stages of different strains have strain-specific surface ant
C:Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The
C:Superfamily: major merozoite surface antigen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1631/Product: major merozoite surface antigen #status predicted <MAT>
F:67-84/Region: 3-residue repeats (S-G-T/P)
F:1614-1631/Domain: membrane anchor #status predicted <MBN>
F:97,259,755,835,911,955,1049,1156,1165,1436,1563/Binding site: carbohydrate (asn

Query Match 94.6%; Score 7965; DB 1; Length 1631;
Best local similarity 95.3%; Pred. No. 1e-239;
Matches 1563; Conservative 19; Mismatches 48; Indels 10; Gaps 3;
QY 1 MKLIFLCSEFLFTINOCYTHESYOELVKLELEDAVLTGYSLFQKEMVLENGSGT 60
D 1 MKLIFLCSEFLFTINOCYTHESYOELVKLELEDAVLTGYSLFQKEMVLENGSGT 60
QY 61 AVTSTPGSGSVASGSGSVASGSGSVASGSGSVASGSGSVASGSGSRTPNSDSS 120
D 61 KGAQAQSGTSGT--SGTSGPGSGTSPSSRSNTLPSSNMSGAS-----PPAADS 111
QY 121 DAKSYADIKHRVNYLTITKELKYPQLFDLTNHLTLCDNHGKRYLIDGYEETINELLYK 180
D 112 DAKSYADIKHRVNYLTITKELKYPQLFDLTNHLTLCDNHGKRYLIDGYEETINELLYK 171
QY 181 LNFYFDLLRAKLVNVCANDVCQIFPNKIKANELDVLAKLVFGYRKPLDNIKDVGMED 240
D 172 LNFYFDLLRAKLVNVCANDVCQIFPNKIKANELDVLAKLVFGYRKPLDNIKDVGMED 231
QY 241 YIKKNNKTIEINELIEESKKTIDKNNKATKEEKKKLYOAOYDLSYNNKOLEAHNLIS 300
D 232 YIKKNNKTIEINELIEESKKTIDKNNKATKEEKKKLYOAOYDLSYNNKOLEAHNLIS 291
QY 301 VLEKRIDTLKKNENIKELDLKINEIKNPANSGNTPNTLLDKKKITEEHEKEITAKT 360
D 292 VLEKRIDTLKKNENIKELDLKINEIKNPANSGNTPNTLLDKKKITEEHEKEITAKT 351
QY 361 IKFNIDSLFNDPLEEYLLLEKKNNDISAKVETKESTENEPNGVYTLSTNDINNAL 420
D 352 IKFNIDSLFNDPLEEYLLLEKKNNDISAKVETKESTENEPNGVYTLSTNDINNAL 411
QY 421 NELNSFGDLINPDYTKEPKSNITYTDNERKKFINELKEKIKIEKKKTESDKSYEDRSK 480
D 412 NELNSFGDLINPDYTKEPKSNITYTDNERKKFINELKEKIKIEKKKTESDKSYEDRSK 471
QY 481 LNDITREYKELINEIYDSKFNNNIDLTNFEKMGKRSYKVEKLTHTNFASYSKHNL 540
D 472 LNDITREYKELINEIYDSKFNNNIDLTNFEKMGKRSYKVEKLTHTNFASYSKHNL 531
QY 541 EKLTKALKMEDYSLRNIVVEKELKYKKNLSKISKENIETLVENIKKDEDLFEKKITKD 600

```
Db 532 EKLTKALYMEDYSLRNTVEKELKYKNLSKLENEIETLVENIKKDEQDFEKKIKD 591
Qy 601 ENKPEKILEVSDIYKVOVKVLMNKIDELKKTQILKKNVELKNHVPNSKQENKOE 660
Db 592 ENKPEKILEVSDIYKVOVKVLMNKIDELKKTQILKKNVELKNHVPNSKQENKOE 651
Qy 661 PYLLIVLKEIDKLVKPMPEVSLINEBKKNITKQGSQDNSEPTTEGELTGOATTPCQO 720
Db 652 PYLLIVLKEIDKLVKPMPEVSLINEBKKNITKQGSQDNSEPTTEGELTGOATTPCQO 711
Qy 721 AGSALGSGVQAQOEOQAOQPPVPPVPEAKAQPPTPPAPVNNKTEVANSKIDYLEKLYE 780
Db 712 AGSALGSGVQAQOEOQAOQPPVPPVPEAKAQPPTPPAPVNNKTEVANSKIDYLEKLYE 771
Qy 781 FLNTSYICHKTIIVSHSTPMNEKIKQYKITKEEESKLSQCDPLDLFNIONNIPVPMF 840
Db 772 FLNTSYICHKTIIVSHSTPMNEKIKQYKITKEEESKLSQCDPLDLFNIONNIPVPMF 831
Qy 841 DSLNLSLSQLEMEIYKEMVCLYKLDNDKIKNLLEBAKAVSTSVKTLSSSMQPLSLT 900
Db 832 DSLNLSLSQLEMEIYKEMVCLYKLDNDKIKNLLEBAKAVSTSVKTLSSSMQPLSLT 891
Qy 901 PODKPEVANDTSHSTLNSLSIKFENILSLGKNKNITQELIGKSSSENEYEKILKQSD 960
Db 892 PODKPEVANDTSHSTLNSLSIKFENILSLGKNKNITQELIGKSSSENEYEKILKQSD 951
Qy 961 TFYNESEFNFVKSADINSINDESKRKLLEEDINKLTKTQLSFDLNNKKLKERLFD 1020
Db 952 TFYNESEFNFVKSADINSINDESKRKLLEEDINKLTKTQLSFDLNNKKLKERLFD 1011
Qy 1021 KKKVGVGKWKQIKKLTLEKQESKLSLNNPKHVLQNFVFNKKKAEIETENTLEN 1080
Db 1012 KKKVGVGKWKQIKKLTLEKQESKLSLNNPKHVLQNFVFNKKKAEIETENTLEN 1071
Qy 1081 TKILKHKGVKVKYNGSSSLKLTSEESIOTEDQVYASLNFVKYLSKLEGLKQDLNLEK 1140
Db 1072 TKILKHKGVKVKYNGSSSLKLTSEESIOTEDQVYASLNFVKYLSKLEGLKQDLNLEK 1131
Qy 1141 KKLVSLSGLLHLIAELKEVINKNNKYTNSPSENNTDVNNKLESKFKLPCTGDAVAVS 1200
Db 1132 KKLVSLSGLLHLIAELKEVINKNNKYTNSPSENNTDVNNKLESKFKLPCTGDAVAVS 1191
Qy 1201 EGGSDTLEQSPKRPASTHVGAESENITTSQNVDEVDVITVPIFGSEEDYDGLQGV 1260
Db 1192 EGGSDTLEQSPKRPASTHVGAESENITTSQNVDEVDVITVPIFGSEEDYDGLQGV 1251
Qy 1261 TGEAVTPSVINILSKITENEVEVLYLKLPLAGVYSLSKQLENNVWTFVANNKQILNSRFN 1320
Db 1252 TGEAVTPSVINILSKITENEVEVLYLKLPLAGVYSLSKQLENNVWTFVANNKQILNSRFN 1311
Qy 1321 KRENFKNVLESDELTPYKDLTSSNVVADPYKFLNKEKDRKFLSSYNTKQSDIDTDINFRAN 1380
Db 1312 KRENFKNVLESDELTPYKDLTSSNVVADPYKFLNKEKDRKFLSSYNTKQSDIDTDINFRAN 1371
Qy 1381 DVLGYKYLSEKYSKSDLSIKKYINDKQGENEKYLPFLNNIETLYKYVNDKIDLFVJHLE 1440
Db 1372 DVLGYKYLSEKYSKSDLSIKKYINDKQGENEKYLPFLNNIETLYKYVNDKIDLFVJHLE 1431
Qy 1441 AKVANTYKESKNSVEKIKELNYLKTIDQKLDLDFKKNNFVGIADLSDYNNINNLTKFLS 1500
Db 1432 AKVANTYKESKNSVEKIKELNYLKTIDQKLDLDFKKNNFVGIADLSDYNNINNLTKFLS 1491
Qy 1501 TGVAFENILAKVNLNLGDNLT-OGMLNLSHQCVKQCPONGSGCFRHLDEREECKCLNTY 1559
Db 1492 TGVAFENILAKVNLNLGDNLT-OGMLNLSHQCVKQCPONGSGCFRHLDEREECKCLNTY 1551
Qy 1560 KOEGDKCVENPPTCENNGGCDADAKTEBDSGNGKIKCECTKPDYSYLFDFGIFCSS 1619
Db 1552 KOEGDKCVENPPTCENNGGCDADAKTEBDSGNGKIKCECTKPDYSYLFDFGIFCSS 1611
Qy 1620 SNFLGISFLILMLILYSFI 1639
Db 1611 SNFLGISFLILMLILYSFI 1601
```

```
Db 1612 SNFLGISFLILMLILYSFI 1631
RESULT 4
SAZQM
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
N:Alternate names: 195k glycoprotein
C:Species: Plasmodium falciparum
C:Date: 30-Sep-1987 #sequence_revision 31-Mar-1991 #text_change 09-Jun-2000
C:Accession: A23386; S06361
R:Weber, J.L., Leininger, W.M., Lyon, J.A.
Nucleic Acids Res. 14, 311-323, 1986
A:Title: Variation in the gene encoding a major merozoite surface antigen of the huma
A:Reference number: A23386; MUID:86205236
A:Accession: A23386
A:Molecule type: DNA
A:Residues: 1-1104 <WEB1>
A:Cross-references: EMBL:X03831
R:Weber, J.L., Sim, B.K.L., Lyon, J.A., Wolff, R.
Nucleic Acids Res. 16, 1206, 1988
A:Title: Merozoite surface protein sequence from the Camp strain of the human m
A:Reference number: S06361; MUID:8814399
A:Accession: S06361
A:Molecule type: DNA
A:Residues: 1104-1726 <WEB2>
A:Cross-references: EMBL:X03831
C:Comment: The merozoite stages of different strains have strain-specific surface ant
C:Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat
F:1-19/Domain: signal sequence #status predicted <Sig>
F:20-176/Product: major merozoite surface antigen #status predicted <Mat>
F:67-87, 91-96, 100-105, 109-120/Region: 3-residue repeats (S-G-T)
F:757-765/Region: 3-residue repeats (T-E-E)
F:133, 272, 501, 567, 638, 827, 839, 924, 944, 990, 1016, 1114, 1221, 1613, 1658/Binding site: carb
```

Query Match 61.0%; Score 5136.5; DB 1; Length 1726;
Best Local Similarity 60.1%; Pred. No. 4.1e-152;
Matches 1052; Conservative 221; Mismatches 342; Indels 135; Gaps 26;

```
Qy 1 KMTIFPLCSLFTINTQCVTHESYQELVKLEALDVAULTGYSLFOKEKMYLNE---- 55
Db 1 KMTIFPLCSLFTINTQCVTHESYQELVKLEALDVAULTGYSLFOKEKMYLNEEITTT 60
Qy 56 -----GTSQAVTTSPPGSKGSVAGSGSGSVAGSGSVAGSGSVAGSGSVAGSGS 105
Db 61 KGASQSGTSGTSGTSGTSGTSGTSGTSGTSGTSGTSGTSGTSGTSGTSGTSGTSGT 118
Qy 106 GN--SRRTN-----PSDNSSPSDAKSYADLKHVRNLTITKELKYPQLFDLT 151
Db 119 GTSPPSSSNTLPRSNSTSSGASPPADASDSQKSYADLKHVRNLTITKELKYPQLFDLT 176
Qy 152 NHTLTLDNTHGFKYLLIDGVEINELLYKLNIFYDILLAKLINDVCANDYCOIPNFKIRA 211
Db 179 NHTLTLDNTHGFKYLLIDGVEINELLYKLNIFYDILLAKLINDVCANDYCOIPNFKIRA 238
Qy 212 NELDVLLKLVGKYPKPLDNIKDNNGKMEDYTKKKKKTINENINELIESKTIQKNKATK 271
Db 239 NELDVLLKLVGKYPKPLDNIKDNNGKMEDYTKKKKKTIANINELIESKTIQKNKADN 298
Qy 272 EEEKKKLYQAOYDLSINLKOLEFAHNLISVLEKRIDPLKKNENIKELLDKINETKPNPPA 331
Db 299 EEGKKKLYQAOYDLSINLKOLEFAHNLISVLEKRIDPLKKNENIKELLDKINETKPNPPA 358
Qy 332 NSGNTPTLLDKNNKIEHEKEIKELAKTIKENIDSLFTDPLELEYLVRKNNKIDISAK 391
Db 359 NSGNTPTLLDKNNKIEHEKEIKELAKTIKENIDSLFTDPLELEYLVRKNNKIDVTPK 418
Qy 392 VE--TKESTEEN--EYPMGVATYPLSYNDINNAL--NELNSFGDILNFDYTKESKAIYT 445
Db 419 SODETKSVQIIPKVPYPIGVIYPLPLTDIHNLSLADNDKNSYGDLMNP--DKREKINEKIT 477
```

Oy	446	DN-ERKFEINIEKIKIEKKI-----EDKSYEDRSUMDITKEVEKLLINLEYDSKF	500
Db	478	DNKRKLFINNKKQIDLEKKIKNTHREONKKLEEDYKS----KKDYELLEKEEKKF	533
Oy	501	NNNIDLNFPEKMMGKRXYVEKLTTHNPFASYENSKHNLEKLTAKLYMEDYSLRNTIV	560
Db	534	NNNFKDROVDKISARTYVEKORNNKPFSSNNSVYNQKALKALSLYEDYSLRKIS	593
Oy	561	EKELKYKNLISKIENEIETLVENIKKDEODLEKK--ITKDEKNPDEKILEVSDIVKY	617
Db	594	EKDENNHYTLTGLEADIKKLTPEIKSSENKILEKNFKGLTHSANMS---LEVYDIYKL	649
Oy	618	QOVQVLLMMKIDELKTQULIKVLEKHNHVPNSYKQENKQEPYLYLVKKEIDKLVF	677
Db	650	QOVQVLLIKRIEDLRKLEIFELKNAQKDSIHVPNIYKPPKNDEPYLLYLVKEDDKLEF	709
Oy	678	MPKVESLINEKKNIKTEEGSDNSEPSTGEITGATTPGQQAQASALEGDSVQAQAQO	737
Db	710	IPKPKDMLKKEQAVLS-----ITQPLVAASSTTEDGSHHTTISQSGEVEYEEET	760
Oy	738	KQAOAPV-----RPVPEAKAOVPTPEAPVNNKTEVMSKLDYLEKLEFLNTSYCH	789
Db	761	EETBEVYGHNTTYITLTPREVAUVENSLEHKSNDNSQALTYVTYLLKLDLEFLTSYICH	820
Oy	790	KYILVSHSTANEKILKOYKLTKEESKSSCORLDDLFFNIQNNIPVMTSMEDSLNNSISQ	849
Db	821	KYILVSSNSMDOKLEEVYNTLPEEENELSCDPLDLFFNIQNNIPAMSLVSDSMNDLQH	880
Oy	850	LFEMEYERKEMVNCYKRYKNDKDIKNLEEKVY-----SPSYVTLSSS-----	893
Db	881	LFELTYKEMITYLHLKKEENHILKLEBOKQITGTSSTSPGNTVTYTAOSATHNSQN	940
Oy	894	--MOPLSLTPOQDREVSAND--DTSHS---TNLNSLSLEFENLISLGGKNKIYOEL-I	943
Db	941	QOSNASSTNQNGAASVSGPAVEESHDPJLVYLSNDLKGIVSLNLTGNKTKVPNPLTI	1000
Oy	944	GQKSENFYERKILKDSPTFNESFTNPKVSKADDINSLDESKRKKLEEDINKLKTQL	1003
Db	1001	STTEMEFERYENILKNDTYTNDIDIKOFVNSNSKVITGLT-EPOKNALEDEIKKLDIQL	1059
Oy	1004	SFDLYNKYKLTLELFPKRRKTVGYKMOJKLTLLEKQESKLSNMPKHVLONFVSVF	1063
Db	1060	SFDLYNNKTKLDRLEFKKKLEQDOKMOJKLTLLEKQESKLSNMPKHVLONFVSVF	1119
Oy	1064	NKKKEAIEAETNLENTKLTILKHYKGLVKKYVNGESSPLKLTISESIOTEDNYASLENFK	1123
Db	1120	NKKKEAIEAETNLENTKLTILKHYKGLVKKYVNGESSPLKLTISEVSIQEDNYANLEKFR	1179
Oy	1124	VLSLEKCLKDNLNEKKKLSYLSGLJHLJAEKLEVINKNKYTGNSPENNTPVNNALE	1183
Db	1180	VLSIKDGLKNDNLHLGKKKLSFSSGLJHLJTELKEVINKNKYTGNSPENNKKVNEALK	1239
Oy	1184	SYKFELEGGDVAFVYSE-----SGSDTLEQSPKKPASTHGAESNTITTT	1229
Db	1240	SYENFLPE-AKVTYVYTPPOPDVTPSPPLSVRSVSGSGSTKEETQIPTSGSLTELOOVQ	1298
Oy	1230	SQNVDDDEDVIIATPIFGSESEEDYDLAGOVVGEATVPSVINIISIKIENEVEYVATPL	1289
Db	1299	LQNVDEDDSLVYLPPIGSESDNDEYLDQVYVGEALSVT-MONILISGENEVDYIYKPL	1357
Oy	1290	AGVYRSILKQOLENNMTFNVNKOILNSRPNKRENFKNVLESDDLPIYKDLTSSNVAKDP	1349
Db	1358	AGVYRSILKQOLENNMTFNVNKOILNSRKLKRYKFLVLEDDLMOFKHISSENEYIEDS	1417
Oy	1350	YKPLNKKRROKPLSSVNYIKDSIDTQIPNANDVLYKYLKSSKYSQSDSISIKKYI----	1404
Db	1418	FKLNSQKNTLKSTKYIKIESVENDIKPAQBSIYERKVLAKYKODLESIKKYIKEKE	1477
Oy	1405	-----NDKQENKRYLPLFNINNIETLKYTVNDRIDFVJHLEKVLANTYVE	1449
Db	1478	KFPSPPTTPPSPAKTDQEKESKFLPLPLTNIETLVNKLVNKIDYLLINKAKKINDCAVE	1537
Oy	1450	KSNEVAKIKELNLYKTQDKLADFPKNNNFVGIADLSYDYNHNNLLTKPLSTGAVFENIA	1509

[illegible]

478 DNKEKRIFFINNKKOIDEKKINHTKEQNKLLDEYKS---KKDYEELEKEFEYENK 533
501 NNNDILTFEKKMKRYKYVEKLEHTHNTFASYENSKHNLEKLTALKYMEDYSLRNTIV 560
534 NNNDKDVVDKIFISARYYNNKORNNKPFSSNNNSYVNNVOKLAKALSYLEDYSLRKIS 593
561 EKEKYYKNLISKIENELETLEVENIKKDEQLFEKK---ITKDNKPKDEKILEVSDIVK 617
594 EKDNNHYTLEKGTLEADIKLLEEKISSENKILEKFNKGLTHSANAS---LEVYDIKVL 649
618 OVQVLLMKNKIDELKKTOLLKNEELKHNHVPNSYKQDNQOEPRYLLVKKEDKLKVF 677
650 OVQVLLKIKEDIKLEKLEFKNAOLKDSIHVPNLYKPPONPEPYLLVLYKEDVLIKLEF 709
678 MPKESLITNEKKNKIKTEGQSDNSPTEGELTGQATKKPCQAGSALEGDSVQAQAOEQ 737
710 IPKVKMKKQOAVLS-----ITQPLVASSETTEDGSHNTLSOGSEFEVTEET 760
738 KOAOPV-----PVPVPEAKAOPVPPAPVNNKTEVNSKLDYLEFELNTSYICH 789
761 EETEBVTGHTTTVTITLPRKEVKYVENSIKHSNDNSQALIKTYLKLDELFTKSYICH 820
790 KYIIVSHSTAMEKILKOKYKITEESKLSGCDPLDLFNIONNIPVMYSMFDLSNLSIQ 849
821 KYIIVSNSSMOQKLELVNLTPEEENELKSCDPLDLFNIONNIPVMYSLYDSMNDLQH 880
850 LFMETYEKEMVKNLYKLDNDKIKNLBEAKV-----STVYKTLSSSS----- 893
881 LFFELYQKEMYYIKLKEENHIKKLEBQKOITGSTSTSPGNTVTVTAOSATHSNQON 940
894 --MOPSLTPDKREVSAND--DTSHS---TNLNSLKLFEENISLGNKNKIYQEL-I 943
941 QOSNASTNTONGVAVSSGPAVVESSHDLPLVLTISNDLKGIVSLNNGKTKVNPPLTI 1000
944 GOKSENFYEKILDSDFEYESFNFYKSKADDINSLDSKSKKLEEDINLKLTIQL 1003
1001 STTEMKEFYEILKNNDYFENDDIKQFVKSNSKYITGLT-ETOKNALDELKIKLDTQL 1059
1004 SFDLYNKKYKLERLEFDCKTVGKYKMOIKLTLLEKLESKLSNLPKRVLONFVSEF 1063
1060 SFDLYNKKYKLERLEFDCKTVGKYKMOIKLTLLEKLESKLSNLPKRVLONFVSEF 1119
1064 NKKKEAEIETENTLENTKILKHYKGLVKKYNGESSPLKTLSESIOTEDNVASLENK 1123
1120 NKKKEAEIETENTLENTKILKHYKGLVKKYNGESSPLKTLSESIOTEDNVANLEKFR 1179
1124 VLSLLEGLKLNULLEKKKLSYSSGLHLLAELKEVYKKNNTYNGSSENNVDYNNALE 1183
1180 VLSKIDGLKLNULHGLKKKLSYSSGLHLLAELKEVYKKNNTYNGSSENNKKNVNEALK 1239
1184 SYKFFLEPGTVAIVVSE-----SGSDTLEQSQPKKPASTHGAESNTITTT 1229
1240 SYENFLPE-AKVTVTVVTPPOQDVTTPSPLSVAVSSGSGSTKETQIPISGSLTLEQOYVQ 1288
1230 SQNVADVEDVIVIPFGESEEDYDGLQVVTGAVTPSVIDNLSKIENEYVLYLKLPL 1289
1299 LQNYDEEDDSLVLPFICESEDNEDYDQVVTGAVT-MDNLSGSENEYDVIYLLKPL 1357
1290 AGVYRSJLKOENNVMTFNVAVKDLNSRFKKNREKFAVLESOLTPYKDLTSSNYVAKDP 1349
1358 AGVYRSJLKOENNVMTFNVAVKDLNSRFKKNREKFAVLESOLTPYKDLTSSNYVAKDP 1417
1350 YKFNKKEKROKFLSSYVNIKSDIDINFANDVGYKYLISEKYSKSDSISKKYI----- 1404
1418 FKLNSQKNTLKSYYKIKSYKISVENDIKFAOEGISYIEKVLAKKDLSEISKKYIKEKE 1477
1405 -----NDQGENEKYLLPLNIEFLYKTVNDKIDLEFVTHLEKAVLNTYE 1449
1478 KFPSSPTTPPSPYKTDQKKESKFLPLFLNIEFLYNNLVKNDYILNIAKALINDCAVE 1537
1450 KSNVEVATKELNTYKLTQDKLADFKKNNNPGVADLSIDYVHNNLTKFLSTGAVFELA 1509
1538 KDEAHVATITKLSLDAIDKIDLFKNNHNDFAIKLINDTKKMDLGLSTGLV-QNPF 1596

1510 KTVLSNLDGNLQGMNLNISQHCYKQOPONSQCFRHLDEBECKCLNTYKQEGDKVEN 1569
1597 NTIISKILEGKFOMLNLSQHCYKQOPENSQCFRHLDEBECKCLNTYKQEGDKVEN 1656
1570 PNPCTENNNGCCDADAKTEEDSGSNKKITCECTKPDSPYLPFGJICSSNPLGISFLL 1629
1657 PNPCTENNNGCCDADAKTEEDSGSNKKITCECTKPDSPYLPFGJICSSNPLGISFLL 1716
1630 LTMILYSFI 1639
1717 LTMILYSFI 1726
RESULT 6
A54498
major: merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C: Species: Plasmodium falciparum
C: Date: 28-Oct-1994 #sequence: revision 28-Oct-1994 #text_change 09-Jun-2000
C: Accession: A54498
R: Peterson, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brow
Mol. Biochem. Parasitol. 27, 291-302, 1988
A: Title: Variation in the precursor to the major merozoite surface antigens of Plasm
A: Reference number: A54498; MUID:88142999
A: Accession: A54498
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-1701 <PEP>
A: Cross-references: GB:M19143; NID:g160412; PIDN:AAA29653.1; PID:g160413
C: Superfamily: major merozoite surface antigen
C: Keywords: surface antigen
Query Match 60.18; Score 5065; DB 2; Length 1701;
Best Local Similarity 60.18; Pred. No. 6; 6e-150;
Matches 1042; Conservative 227; Mismatches 338; Indels 126; Gaps 29;
1 MKIIFLCSPFLFTINQCVTHESYQELVKLEALEDAVLVGYSLFQKKNVLENGTSGT 60
1 MKIIFLCSPFLFTINQCVTHESYQELVKLEALEDAVLVGYSLFQKKNVLENGTSGT 60
61 AVTSTPGSGSVASGSGSVASGSGSVASGSGSVASGSGSGNSRTPNSDSSS 120
61 AVTSTPGSGSVASGSGSVASGSGSVASGSGSVASGSGSGNSRTPNSDSSS 114
121 DAKSYADLKHVRNLYLTIELKLPOLFDLTNMLTCLDNIHGKYYLIDGEEFINELLYK 180
115 NFKTYADLKHVRNLYLTIELKLPOLFDLTNMLTCLDNIHGKYYLIDGEEFINELLYK 174
181 LNFYFDLRAKLNACANDYQOIFPNLIRANEDVLKLELVGYRKPDLNIDKQVGMED 240
175 LNFYFDLRAKLNACANDYQOIFPNLIRANEDVLKLELVGYRKPDLNIDKQVGMED 234
241 YIKNNKTTIENINLEBSKKTIDKNKNATKEEKKKLYQAQDLSYNNKOLEAHNLIS 300
235 YIKNNKTTIENINLEBSKKTIDKNKNADNEBKKKLYQAQDLSYNNKOLEAHNLIS 294
301 VLEKRIDTLKKNENIKELDLKINIK-----NPPRANGNPNTLLD-KNKKIEHEKEIK 355
295 VLEKRIDTLKKNENIKELDLKIDKIDAEKP-----TTGSKRNPPLPENKKKEVEHEBEKIK 351
356 EIATIKFNIDSLFTDPLELEYLYREKNNKIDISAKYE--TKRESTEEN-EVNPQVTPPLS 412
352 EIATIKFNIDSLFTDPLELEYLYREKNNKIDISAKYE--TKRESTEEN-EVNPQVTPPLS 411
413 YNDINNAL--NELNSFGDLINPEDYTKEPSKNITYDN-ERKKEFINIEKIKIEKKKI- 467
412 LTIHNSLADNDKNSYGDILMP-DTEKINEKILITDNKKEKIFINNKKOIDEERKIN 470
468 ---ESDKKSYEDRSKSLNDITKEYEKLLNELYDSKFNNNIDLNFEEKMGKRSYKYEKL 524
471 HFKQONKKLLEDEYKS---KKDYEELEKEFEYENKFNNDKDVVDKIFISARYYVNEKQ 526

Oy	525	THNHTFASYNESKNIETLKRLKALKWEDSLRNIYVEKELKYXKNLSIKENIEIETLVEN	584
Db	527	RYNKFSSNSNYSVNOVKLALKSTALYLEDVSLRGISEKOPNHYTLTKGLELKTGELKEE	586
Oy	585	IKKDEOLFEKK---IYKDNKSPDEKILEVSDIVKVOVOYKVLMMNKIDELKKTOLILKNV	641
Db	587	IKSENNKILENFKGLHSHANAS----LEVSDIVKIOVOYKVLILKIDELKRIELFLKNA	642
Oy	642	ELKNNHVPNSYKOEKOEPEYLIYVLKKEIDKLVMPVESLINEEKNNIKTEGOSDNS	701
Db	643	OLKSHIHVPNYKOPKPEPYLIYVLKKEVDKLEKPIPKYMDLKKEGQVLISS-----	695
Oy	702	EPSTEGELTGATTKPEGOAGSALDEDSVOAOQOEOKA-----QPPYPRVPEAKAQ	754
Db	696	--TIQPLVASETTEDEGSHSTHTL--SQSGETEVEETEVEETVEGHTTWTVTITLPEKES	752
Oy	755	VPTPAPVNNKNTEN-----VSKIDLYLEKLYEAFNTSYICHKYILVSHSTPMNEKILQ	806
Db	753	APKEVYKVENSIERKSDNSQALTITKYVLKKEIDFLTKYICHTIKYILVNSNMDKLLVY	812
Oy	807	YKTKKEBSKILSCDPLDLFTNIONNI PWAYSMFDSLNSLSQLFMEIYERKEMVCNLYKL	866
Db	813	YNLPEEENELKSCDPLDLFTNIONNI PWAYSLYDMSNIDLHTEFELYOKEMIYYLHKL	872
Oy	867	KDNOKINLLEAKKV-----STYKTLSSSS-----MQPLSLTPQOKPEVS	908
Db	873	KEENHRIKLEBOKOILTGTSSTSPGGTWTYVNAQSYTHNSQNOOSNASTTQOGVANS	932
Oy	909	AND---PTSHS---TNLNSLSLFTENISLGKNNKIYOEL--IGORSSEPFERYELKSD	960
Db	933	SGPAVVEESHDPILTVLISNDLKGIVSLNKGKTKYVPPRLITSTTEMEKFTENILKND	992
Oy	961	TFVNESTNPEYKSKADDINSLNDESKRKLEKIDNKLKTQLOSPDYKNKYKLERLFD	1020
Db	993	TYFNDDILKOFEYKSNMSKYITGLT--ETQKNALNDEIRKIKDTLOLSPDYKNKYKLDRLFN	1051
Oy	1021	KKKRVGKYNKQIKKTLTKLQOLSKLNSLANKPNHYLONSVFPNKKKAEIATETNTLEN	1080
Db	1052	KKKELGODKQIKKTLTKLQOLSKLNSLANKPNHYLONSVFPNKKKAEIATETNTLEN	1111
Oy	1081	TKILKHYKGLVYKYNESSPKLTLSSESIOTEDBNYASLENEKRVLSKLEGLKJNDMLNK	1140
Db	1112	TKILKHYKGLVYKYNESSPKLTLSSESIOTEDBNYANLEKRALSKIDGKJNDMLHKG	1171
Oy	1141	KKLISYLSGHLHLIAELKEVIKKNKNTGNSPSENNTDVNNALSKYKPLPECTDYATVVS	1200
Db	1172	KKLISYLSGHLHLITELKEVIKKNKNTGNSPBNKKVMEALKSYENFLPE--AKTVYVT	1230
Oy	1201	E-----SSDPLTEOSQPKRKPASTHVGAEKSNFTITTSQNDVEDVDYVITPFL	1246
Db	1231	PPQDVTYPPSPILSVKSVSSSGSTKEETQOIPSGILTEILOOVOLYNDEEDSLVLPFLF	1290
Oy	1247	GESEDEYDDLGQVVTGAVGPVYIDNLSIKTENEBYEVULKPLAGYYSIRSKOLENNWTF	1306
Db	1291	GESEDNDEYLDQVYTGGAISVTF--MDNLSGFENEBYVITKPLAGYYSIRSKOLEKNTIT	1349
Oy	1307	FNVNVKDLNSRFNKRENFKNVLESOLIPYKDLTSSNYVYKPYFLNKEKRDKLSYNS	1366
Db	1350	FNLMLNDILNSRLKRRKYFLVDVLESJDMQFKHTSSNEYIIEEDSFILNSSEQNTLKSXYK	1409
Oy	1367	YIKRISITDINFANDVLGYIKLISEYKSDLSIKKYT-----ND	1406
Db	1410	YIKSESVENDIKFAOEGISYEVKVLAKYDLEIKVILEEKKEKPPSPBPTPPSPAKTD	1469
Oy	1407	KOGENEKRYLPEFLNNIEFLYKTVMDKIDLPYHLEAKVLVYTYEKSVAEVIKELMYLAKTI	1466
Db	1470	EOKESKFLPELITNIEFLVNLNVLNKIDTVLINKAKINCANVEKDBAHKTKTSLSDLAI	1529
Oy	1467	QDKLADRRKNNNFVGIADLSTDVNNHNNLLTKFLSITGMVNEENLAKTVLSLNDLGNLOQM	1526
Db	1530	DDKIDFLKRNINDFEAIKKLLINDTKKMDMLGKLLSTGLV--QNPNNITISKLIEGKRQMDYLN	1588
Oy	1537	ISOHCYKAKOCPONSGCFRHLDEREECKCLANTYKOEGBKCVENPMTCONENNNGCCADAK	1566

QY	1587	CTEEDSGSGNGKRTCECTKRPDYPLEFDGTCFSSNFIQLISFLILMLILYSFI	1639
Db	1589	ISOHCYVKQCPENSGGCFRHLDERECCCLNTAKOEGKCYENPPTCNENNGCCDADAT	1648
QY	1587	CTEEDSGSGNGKRTCECTKRPDYPLEFDGTCFSSNFIQLISFLILMLILYSFI	1639
Db	1649	CTEEDSGSGRRKRTCECTKRPDYPLEFDGTCFSSNFIQLISFLILMLILYSFI	1701
RESULT	7		
	A26868		
	major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)		
	C.Species: Plasmodium falciparum		
	C.Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 09-Jun-2000		
	C.Accession: A26868		
	R.Tanabe, K.; MacKay, M.; Goman, M.; Scaife, J.G.		
	J. Mol. Biol. 195, 273-287, 1987		
	A.Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmod		
	A.Reference number: A26868; MUID:88011243		
	A.Molecule type: DNA		
	A.Residues: 1-1701 <TAN>		
	C.Superfamily: major merozoite surface antigen		
	C.Keywords: surface antigen		
	F.1-19/Domain: signal sequence #status predicted <SIG>		
	F.20-1701/Product: major merozoite surface antigen #status predicted <MAT>		
Query Match	60.0%;	Score 5053; DB 2; Length 1701;	
Best Local Similarity	60.1%;	Pred. No.1.6e-149;	
Matches 1042;	Conservative 225;	Mismatches 338; Indels 130; Gaps 30;	
QY	1	MKIIFFLCSFLFFINTQCVTHESYOELVKKLEALDAVLGTYSLPQEKVNLBEGTSGT	60
Db	1	MKIIFFLCSFLFFINTQCVTHESYOELVKKLEALDAVLGTYSLPQEKVNLBEGTSGT	60
QY	61	ATTSTSPGSGSVASGSGSVASGSGSVASGSGSVASGSGSGNSRTRNPSDNSDS	120
Db	61	ATTSTSPGSGSVT---SGGVSASVASVASG---SGGVSASGSGNSRTRNPSDNSDS	114
QY	121	DAKSADADLKHRYRNLTLTIKELYPQLFDLNNHMLTLCNDJHGEVLYLDGEEELNELLYK	180
Db	115	NKTYADADLKHRYRNLTLTIKELYPQLFDLNNHMLTLCNDJHGEVLYLDGEEELNELLYK	174
QY	181	LNPFYDILLRAKINDVCANDYCOIPENLKRANELDVLKLVGYRRPLDNIKDVNGKMD	240
Db	175	LNPFYDILLRAKINDVCANDYCOIPENLKRANELDVLKLVGYRRPLDNIKDVNGKMD	234
QY	241	YTKKKKKTENINELIESKRTIDKNNKATYDEEKKKIYQOYDLSTYNNKOLEAHNLIS	300
Db	235	YTKKKKKTIANINELIESKRTIDKNNKADNEEGKKIYQOYDLSTYNNKOLEAHNLIS	294
QY	301	VLEKRIIDLTKKNEIKELLDIKINEIK---NPPRANSGNPTNLLD-KNKKIEEHEKIK	355
Db	295	VLEKRIIDLTKKNEIKELLDIKIKTDAENP---TTSKKNPLPENKKKVEEGHEKIK	351
QY	356	ETAKTIKENIDSLFTDPLELEYLYREKKNKIDISAKVE--TKESTEEN-EYPNGVTPLS	412
Db	352	ETAKTIKENIDSLFTDPLELEYLYREKKNKIDYVPRKSDPTKSYQIVPVPNGVTPLP	411
QY	413	YNDINNAL---NELNSFDLNPFDYTEPSKNITDN-EKKKTIENIEKIKIEKKKI-	467
Db	412	LTDINNSLAADNDKKSYPDLMNP-DTEKEIKENIITDKEKKEIINNIIKKOIDEERNIN	470
QY	468	---ESDCKSYEDRSKSLNDITKEYEKILNELYDSGNFNNDILTNPEKKMGKRYSYKVKL	524
Db	471	HTKEONKKLLEDEYS---KKDIEELLEKEYTEEMKFNPNPKDVAVDKFSARYYNVEKQ	526
QY	525	THHNPFASYENSKHLEKLTAKLKYMEDYSLRNIVVEKELYYUNLISKIENIETETVEN	584
Db	527	RNNKFNSSSNVYUQKTKALSTLEDYSLRKGISGKDFNHYTATKGLADIKKLTVE	586
QY	585	IKKDEQLPEKK---ITYDENKPRDEKILLESVDIYKVOYQKYLNNKIDELKLTOLLKYN	641

```

Db 587 IKSSSENIKEKFKGLTHSANAS----LEVSIDIYKQVLLIKIEDLKIELEFLKNA 642
Qy 642 ELKNIHVNSYKQKNOEYUULVILKKEIDKLVEMKVSLEINEEKNKITEGOSNS 701
Db 643 QLKDSIHVPNLYKQKNEPYUULVILKKEVDKLEFIPKVDMLKEQAVLSS----- 695
Qy 702 EPSEGETIGATKRPQOAGSALEGSVQAOEOKA-----OPPVAPVPEAKQ 754
Db 696 --IIOPLVASETTEEDGSHHTL--SOSGEFEVIEFEFVEEYGHHTTVITLPRKEES 752
Qy 755 VPPRPAPVANKTEN-----VSKDYLEKYLEPLNTSYCHKIYVSHSTMEKILKQ 806
Db 753 APKEVKYVENSIEHKSNDNSQALTKTYVKLDEFLTYSYCHKIYIIVSNSMOQKLEEV 812
Qy 807 YKIKKEESKLSGCDPLDLFNIONNIPVMSMEDSLNNSLSQFMELYEKEMCNLYKL 866
Db 813 YNLPRKEEKLEKSCDPLDLFNIONNIPAMYSLYDSMNNDLOHLFEELYOKEMYYLAKL 872
Qy 867 KDNDRKILKEAKKV-----STSVKTLSSSS-----MOPLSLTPQDKPEVS 908
Db 873 KEENHIKKLEEQKOITGSTSSSPGNTVTNTAOSATISNSQONASSTNTONGAVAS 932
Qy 909 AND---DTSHS---TNLNSKLLENILSGKNKNYOEI--IQKSSSENYEKILKSD 960
Db 933 SGPAVVEESHDPDLTVLSTISNDLKGIVSLNLTGNKTKVPNPLTISTTEMEKEFENILKND 992
Qy 961 TFVNESTNFKSKADDIINSDESRRKLEEDINKLKTOLSPDYKXKTLERLFD 1020
Db 993 TYFMDIJKOFYKNSKYITGLT--EQKALNDELTKLDTLOLSDLNKKKLLDRLEN 1051
Qy 1021 KKKTVGYKMOIKKLTLLKEOLESKLNSLNPKHYLONFVSYFNKKKAEAEIETENTLEN 1080
Db 1052 KKKLFGQDKMQIKKLTLLKEOLESKLNSLNPHVLONFVSYFNKKKAEAEIETENTLEN 1111
Qy 1081 TKILLKHYKGVKYKYNESSEPLKLSSEIOTEDNYASLENFKYLSKEGLKONLNEK 1140
Db 1112 TKILLKHYKGVKYKYNESSEPLKLSSEIOTEDNYANLEKFRALSKIDGKLNLDNLHLGK 1171
Qy 1141 KKLVSLSGLLHLLAELEKVKKNKNTYNSPSENNDVNNALLESKKRLEPGTQVATVVS 1200
Db 1172 KKLVSLSGLLHLLAELEKVKKNKNTYNSPSENNDVNNALLESKKRLEPGTQVATVVS 1230
Qy 1201 E-----SSDTELEGSQPKRPASTHVAESNTITTSQNVDEVDYIIVPFI 1246
Db 1231 PPGQDVPSPSLVAVSGSGSGSTKEETQIPTSGSLTLELOQVVOQONDEEDDSLVLPFI 1290
Qy 1247 GSESEDDDLGGVYTGAVYTPSVIDNLSKTENEYEVLYLPLAGVYRSLEKQOLENNMT 1306
Db 1291 GSESEDDDLGGVYTGAVYTPSVIDNLSKTENEYEVLYLPLAGVYRSLEKQOLENNMT 1349
Qy 1307 FNVVAKDILNSRFNKRRENFVNLVLESDDLIPYKDLTSSNVVADPYKFLNKKERDRFLSSYN 1366
Db 1350 FNVVAKDILNSRFNKRRENFVNLVLESDDLIPYKDLTSSNVVADPYKFLNKKERDRFLSSYN 1409
Qy 1367 YIKDSIDTINFANDVLYGYKYLSEKYSKSLDSIKKYI-----ND 1406
Db 1410 YIKDSIDTINFANDVLYGYKYLSEKYSKSLDSIKKYI-----ND 1469
Qy 1407 KQGENEKYLPPLANNIEFLYKLVNDKIDLEVYHLEAKVUNTYEKSNEVYKELNLYKTI 1466
Db 1470 EOKRESFPLFLTIETLILYNNLVNKRIDYLLNLKAKINDCNEVDEAVKTKLSLDAKAI 1529
Qy 1467 ODKLADRRKNNFVGIADLSTDYNNHNLTLFTSTGMY--PENLAKTYVLSLLGNLOGM 1524
Db 1530 DDKLIDLEKNNFVGIADLSTDYNNHNLTLFTSTGMY--PENLAKTYVLSLLGNLOGM 1566
Qy 1525 LNIISOHCQVKKQCPONSQCFRHDEBECCKLNYKQEGDKCVENPNTCENNNGCCAD 1584
Db 1587 LNIISOHCQVKKQCPONSQCFRHDEBECCKLNYKQEGDKCVENPNTCENNNGCCAD 1646
Qy 1585 AKCTEEDSGSKGKITECTKRPDSYPLFDGIFGSSSNFLGISFLILMLIYSFI 1639
Db 1647 AKCTEEDSGSKGKITECTKRPDSYPLFDGIFGSSSNFLGISFLILMLIYSFI 1701

```

```

RESULT 8
A45604
major blood-stage surface antigen Pv200 - Plasmodium vivax.
C:Species: Plasmodium vivax
C:Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 24-Nov-1999
C:Accession: A45604
R:Gibson, H.L.; Tucker, J.E.; Kaslow, D.C.; Kretzli, A.U.; Collins, W.E.; Kieffer, M.C.
Mol. Biochem. Parasitol. 50, 325-333, 1992
A:Title: Structure and expression of the gene for Pv200, a major blood-stage surface
A:Reference number: A45604; M01D:92158013
A:Accession: A45604
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1751 <GIB>
A:Cross-reference: GB:M75674; NID:g160608; PID:g457336
A:Note: sequence extracted from NCBI backbone (NCBITN:83591, NCBIPI:83592)
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query: Match 36.7%; Score 3093.5; DB 2; Length 1751;
Best Local Similarity 38.2%; Pred. No. 7.6e-89;
Matches 704; Conservative 316; Mismatches 517; Indels 307; Gaps 41;

Qy 1 MKIIFELCSFLFIINOCVTHESYQELKLEALDAVLGYSLFQKEKVLNGETSGT 60
Db 1 MKALLPESLPIFYATKQCEET--ESYKOLVAKDLKLEALVVDGYELFHKRL----- 50
Qy 61 AVTSTFGSKGSVASGSGSVASGGSVASGGSVASGSGSGSRTNPSDSSDS 120
Db 51 -----GENDKIKVEFNASANNNN 68
Qy 121 DAKSYADLKHRVNYLYLTKELKLPQLEPDLTNHMLTLCDNHIGKRYLIDGEEINELLYK 180
Db 69 NOVSV--LTSKIRNFKSFELEQIPGHTDLHLRELAVENPGIKYLVESYEFNOQLMHV 126
Qy 181 LNPFEDLRAKLVNVCANDCOIPFNLIKIRANELDVLKLVFGYRKPLDNTKQVGMED 240
Db 127 INFHDLRAKLVNVCANDCOIPFNLIKIRANELDVLKLVFGYRKPLDNTKQVGMED 186
Qy 241 YIKRKKTIENINEL--IEESKK-----TIDKKNATKEEKKKLYQAOYDLSYTKQ 261
Db 187 FYTNKKTITKNIISDLIAENKKRSGHPTTTNGAGTQPANOSIAASSETTQISSGNSG 246
Qy 262 -----TIDKKNATKEEKKKLYQAOYDLSYTKQ 291
Db 247 SSSGSSSSGSSSTGSGSGTGTGQSPPAADASSTNANYEAKKIYQAVYNTIFYTNQ 306
Qy 292 LEEAHNLISVLEKRIIDLKKNENIKELLDKINEIKNPPA---NSGNTPTLLDKNKIE 348
Db 307 LOEQOKLAVIEKRYKYLKEKDKILVLEQVAKKEKLPSPYPTNLTNVAHKAESKIA 366
Qy 349 EHEKEIKETIANTIKFNIDSLFTDPLELEYLYREKKNKI--DISAKVETKESTEPNE----- 402
Db 367 ELEKKIETIAKTVMFDLGLFTTQAELELEYLYREKAKMAGTILIPESKRSAGTPEKTYPTL 426
Qy 403 ---YPCGVYPLSTNDINNMLNELNS---FQDLINPDYKREPSKNI--YDNEKAKKLINE 455
Db 427 KETYPHGISTYALANSIYELIELEKIGSDTEFQDLONP--DDGQPKPKGILLINETKREKLEK 485
Qy 456 IKKIKIEKKIIEEDKSYEDRSKSLMDITREYKELNIEIYDSKFENNIDLTNEKMMKG 515
Db 486 IMNKIKIEEDLPLMKIEYEKIKYVYEAKNVEFRPANHFEALDLTVENKFTDPRKK 545
Qy 516 RYSYKVEKLTHTNHFASENSKNHLTKALKYMEDYSLENIYVEKELKYXKLNLSKIE 575
Db 546 REAYMEKKKIES--CSEYQNSNLIINKLKQLYLDEYLVLRKDLADEIKHFSMEKKK 603
Qy 576 NEITLIVENTIKKDEQLFEKKITDENKPDKEILVSDIYKVOYKVLNKKIDELKTKQ 635
Db 604 SEIYDLAEOETIRKNNKL-----TIENK-----PDFSGVVELQVOKVILIKIEALKNVQ 652

```

```

QY 636 LILKNVELKHNHVPNSYKOEKNOEPYLLYLKKEIDKLKVFMEKVESLJINEEK-----690
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 653 NLKNAVKODLYIPKYKTISEKPEPYLWMLKREIDKLKOFIKISMIATEKNKPTVA 712
QY 691 --NIKTGOS--DNSESTEG-----EITGAQTYPG-----QOAGSA-LE 726
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 713 AADIVAGSOTIRGASSETGTGNTVNAOTAAVVOPOHOVNAATVOTPGTGHQAGGGEAETQ 772
QY 727 GDSVQAOAQOEOKA-----OPPYVPV-PE-AQAQVTPPAPVNNKTEVNSK 771
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 773 TNSVQAOAOVOTPGAGQYASTOTISQAPVTOASPEAPAAPSPPAVAAPATMSK 832
QY 772 LDYELKYEELNYSYICKYILVSHSTWNEKILKOYKITEEESKL--SGCDPLDLFNI 829
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 833 LEYELKLDLFLSAVACHKHIFVNTSMKKELLQYKLNDEQKINIKTDELDLFLNF 892
QY 830 QNNIPVYSMFDSLNSLSOLEMEYEREMVCLYLKLDND-KIKNLEAKK-----VS 883
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 893 QNNIPVYSIDSMNSMLONLIELYOKEMYNIYKKNKDTKIKAFLETLSKAAAPAQ 952
QY 884 TSVKTLSSSSMQPISLT-PQDKPEVSANDDSSHSTNLNSLKLLENISLCKNNIYOEL 942
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 953 SAAPPSQAGTTPYTTAPVTTTTPVTPSPQTSVYTSPPTPQAEBN--RVVGSNEKEPE- 1010
QY 943 IGOKSENFYEKILKDSDFYNESFTNFVSKADDINSLEDSRKKLEEDINKLKTQ 1002
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1011 ADTAQVEKFEYKHLISQIDK-YNDYFOKFLBESQKDEITKM-DETAKALGALIEELKKLQ 1068
QY 1003 LSPDLVKKYKLTLEFLDKKKTGVKYMOKIKLTLKQLESKLSLNNPKHVLQNSVF 1062
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1069 VSLDHGYKYLKLERLKRLKKNKIKNSKDQIKLTLKKNKLEBRQNLNPNSTVLKNTAF 1128
QY 1063 FNKKKEKIAETENTLTKLKHKGVLKYVYNGESSPLKTLSBESIQTDNTASLENF 1122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1129 FNKKREKEKEVENTLKTLELKYKARAKYVYGEPEPLTSLBESQKDEYNLNEKF 1188
QY 1123 KVLSTLKGKLDKNLNEKKKLSYSSGLHLIAELKEVYKKNNTGNSPSENNVDYNAL 1182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1189 RYLSRLGRLKLNLEKLENISYSSGLHHVLTLEKELIKKAKTISGNDHTNIAVAKEAL 1248
QY 1183 ESKYKFLPEGTVAIVVSESGDTELOS-----OPKKPA- 1216
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1249 QAVGELPK-----VTTQEGASTAATFLPTVPSAVPGGLPGAVGPAAGLPPPPAG 1302
QY 1217 ---STHGA-----ESVTTTTSQNVDEVDYITVPIFGESEEDYDILGOVYTGDA---- 1264
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1303 SVPATGGAAGASTEENVAKAQDYADYKVIAPLPLFGNDDGEE--DQVTTGEASEEA 1361
QY 1265 ---VTPSVINILSKIENEYEVLYLKPLAGYRSLLKQLENNVMTFNNVNDILNSRKN 1321
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1362 PEILVPGI-----SDDYVYLKPLAGMYKTIKQLENNHVAFNNTIDMLDSRLKK 1413
QY 1322 RENKNVLESLLPDKLDTSSNVVYKDPYKFLNKRDKFLSSYNYKDSIDTIFINFAND 1381
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1414 RNFLLEVLSNDLNFKSSSSEYITIKDPRKLLDEKKKKLGSYKYGASIDMLATAND 1473
QY 1382 VLGYKTLSEKYSDDLSIK--KYIND-----KQGEN 1411
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1474 GVTYTNMGELYKTHLDGVYTEIKKVEDDIKODEBELKLGNNVNSQDSKKNFEATAKAE 1533
QY 1412 EKYLPFLNNIETLYKTVNDKIDFLVILHLEAKVLNYYTEKSNAVEYKIKELANTLKIQOLA 1471
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1534 EKYLPFLNLSQKEKESLVSKVNTYTDNLKVIYNNCOLEKKAETITVKLQDYNRKDELE 1593
QY 1472 DPKNNNFVGIADLSTVYNNHNLITKFLSTGMPVENLAKTVLSNLDLGNLQGLMINSQHQ 1531
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1594 EYK-----SEKKNEVSKSGLEKLMKSKLKENESKEILLQNLINQQLLMSSSHT 1646
QY 1532 CVKQCGQNSGCFHLDRECKCLNAYKQBGDKCVENPNTCENNNGCCADAKCTEED 1591
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1647 CIDTNVNDNMACYRYLDTGTEEMRCLTFKEBGKCVASNVYTKCDNNGGCAPEACECKMTD 1706

```

```

QY 1592 SGNSKKITCECTKPDYPLFDGIFGSSSNFLGISFLLMLIL 1635
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1707 S-----NKIVCKCTKEGSEPLEGVEGFCSSSSFLSFLMLLFL 1746

RESULT 9
S16752
major merozoite surface antigen MSA-1 - malaria parasite (Plasmodium falciparum) (fra
N:Alternate names: polymorphic schizont antigen p190
C:Species: Plasmodium falciparum
C>Date: 17-Apr-1993 #sequence, revision 17-Apr-1993 #text, change 09-Jun-2000
C:Accession: A44865; S16752
R:Olafsson, P.; Matile, H.; Certa, U.
Exp. Parasitol. 74, 381-389, 1992
A:Title: Plasmodium falciparum: the repetitive MSA-1 surface protein of the RO-71 iso
A:Reference number: A44865; MUID:92275047
A:Accession: A44865
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1086 <OLA>
A:Cross-references: EMBL:X61930
A:Experimental source: isolate RO-71
C:Genetics:
A:Gene: MSA1
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat

Query Match 35.6%; Score 2999.5; DB 2; length 1086;
Best local similarity 57.3%; Pred. No. 3,66-86;
Matches 637; Conservative 147; Mismatches 238; Indels 89; Gaps 22;

QY 1 KMIIFLCSEFLPIINQCVTHESYQELVYKLEALQAVLIGYSLFOKEKWLNF----- 55
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 KMIIFLCSEFLPIINQCVTHESYQELVYKLEALQAVLIGYSLFOKEKWLNFEEITTT 60
QY 56 -----GTSGTAATTPSPGKSVASGSGSVASGSGSVASGSGSVASGSGSVASGSGNS 108
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 KGASQSGTSGTSGTSGTSGTSGT--SSTSGTSGTSGTSGTSGTSGTSGTSGTSGTSS 118
QY 109 RRTNPSDSSSDAKSYADLKHRYRNNYLLTLEKYPQFLDITNIMLYLCONIHGFKYLI 168
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 GAIRPAD-ASDSAKSYADLKHRYRNNYLLTLEKYPQFLDITNIMLYLCONIHGFKYLI 177
QY 169 DGYEINELKLYFPLPAKLNDVCANDYCOIPNMLKIRANLVLKLYGKRP 228
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 178 DGYEINELKLYFPLPAKLNDVCANDYCOIPNMLKIRANLVLKLYGKRP 237
QY 229 DNIDNMGKMDYIKKNTKTENINELIEESKKTIDKKNATKFEKKKLYQAOYDLSIY 288
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 DNIDNMGKMDYIKKNTKTENINELIEESKKTIDKKNATKFEKKKLYQAOYDLSIY 297
QY 289 NKOLEEAAHLISYLEKRIIDLTKKNENIKELDKINEIK-NPPPANSNGTPTLLDKNKKI 347
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 298 NKOLEEAAHLISYLEKRIIDLTKKNENIKELDKINEIKELDKINEIKELDKINEIK 357
QY 348 EEHEKEIKELAKTIFKNIDSLFTDPLELEYTLREKKNKIDISAKVE--TKSTEPN-EYP 404
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 358 EEHEKEIKELAKTIFKNIDSLFTDPLELEYTLREKKNKIDISAKVE--TKSTEPN-EYP 417
QY 405 NGVYPLSYNDINNAL---NELNSFGDLINPEDYTKESKNINLYDN-ERKKEINERKEKI 460
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 418 NGVYPLSYNDINNAL---NELNSFGDLINPEDYTKESKNINLYDN-ERKKEINERKEKI 476
QY 461 KIEKKI-----ESDKSYEDRSKSLNDITKYEKLNLNIDYSKNNNIDLTFEKKMMGR 516
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 477 DLEEKINIHTRQOKKKLEDEYKS---KDYELLEKEFYEMKFFNNNDKDVVVDKIFISAR 532
QY 517 YSYKVEKLTHTNTFASYSNKHNLKELTKALKYMEDYSLRNIVYKELKYKKNLSKLEN 576
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 533 YTVNVEKLIYNNKSSSNYSNVQOKLKAALSYLEDYSLRGKISEKQDNHYYTTLKTGLEA 592
QY 577 EIELVENIKKDEBQLFEKK---ITKDNKPKDEKILEVSDIVKQVAVLLMNNKIDELKK 633

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2001, 13:14:34 : Search time 23.06 seconds
(without alignments)
1463.468 Million cell updates/sec

Title: US-09-269-874-3
Perfect score: 8424
Sequence: 1 MKRIFFLCFLFFIINTQCV.....SNFLGISFLILMLILYSFI 1639

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	615	7.3	115	US-08-152-922A-7	Sequence 7, Appli
2	356.5	4.2	3248	US-08-353-700-1	Sequence 1, Appli
3	356.5	4.2	3248	PCT-US95-16216-1	Sequence 1, Appli
4	351	4.2	1786	US-08-973-462-8	Sequence 8, Appli
5	350.5	4.2	2482	US-08-328-254-6	Sequence 6, Appli
6	332	3.9	976	US-09-104-324B-4	Sequence 4, Appli
7	308	3.7	1312	US-08-687-080-51	Sequence 51, Appli
8	307	3.6	1312	US-08-592-126-148	Sequence 148, App
9	301	3.6	53	US-08-290-919-3	Sequence 3, Appli
10	298.5	3.5	2710	US-08-405-496A-6	Sequence 6, Appli
11	298.5	3.5	2710	US-08-405-496A-6	Sequence 6, Appli
12	298	3.5	3135	US-08-323-170B-2	Sequence 2, Appli
13	296	3.5	1588	PCT-US93-07261-11	Sequence 11, Appli
14	296	3.5	1663	PCT-US93-07261-16	Sequence 16, Appli
15	293.5	3.5	2391	US-08-446-855A-2	Sequence 2, Appli
16	293.5	3.5	2391	US-09-150-741-2	Sequence 44, Appli
17	292.5	3.5	3418	US-08-755-587-44	Sequence 4, Appli
18	288.5	3.4	3418	US-08-603-753D-4	Sequence 4, Appli
19	288.5	3.4	3418	US-09-089-753-4	Sequence 4, Appli
20	288.5	3.4	3418	US-08-986-106-4	Sequence 4, Appli
21	282	3.3	53	US-08-290-919-4	Sequence 4, Appli
22	280	3.3	48	US-08-290-919-12	Sequence 12, Appli
23	279.5	3.3	1589	US-09-356-952-4	Sequence 4, Appli
24	279.5	3.3	2329	US-08-755-587-16	Sequence 16, Appli
25	279.5	3.3	3418	US-08-639-501-2	Sequence 2, Appli
26	279.5	3.3	3418	US-09-044-946-2	Sequence 2, Appli
27	279.5	3.3	3418	US-09-044-908-2	Sequence 2, Appli

28	278.5	3.3	2366	US-08-480-604A-10	Sequence 10, Appli
29	278.5	3.3	2366	US-08-405-496A-10	Sequence 10, Appli
30	275.5	3.3	1093	US-09-315-793-52	Sequence 52, Appli
31	274	3.3	48	US-08-290-919-1	Sequence 1, Appli
32	272	3.3	1354	US-08-685-871-2	Sequence 2, Appli
33	272	3.2	1535	US-08-755-587-185	Sequence 185, Appli
34	271	3.2	48	US-08-290-919-2	Sequence 2, Appli
35	271	3.2	2154	US-08-841-349-4	Sequence 4, Appli
36	270	3.2	2052	US-08-790-912-2	Sequence 2, Appli
37	269	3.2	1435	US-08-568-459A-4	Sequence 4, Appli
38	269	3.2	1435	US-08-487-826B-4	Sequence 4, Appli
39	267.5	3.2	1388	US-08-685-576-4	Sequence 4, Appli
40	266.5	3.2	1579	US-08-755-587-184	Sequence 184, Appli
41	265	3.1	1363	US-08-425-061-23	Sequence 23, Appli
42	265	3.1	1363	US-08-825-886-23	Sequence 23, Appli
43	265	3.1	1852	US-08-425-061-24	Sequence 24, Appli
44	265	3.1	1852	US-08-825-886-24	Sequence 24, Appli
45	265	3.1	1863	US-08-425-061-16	Sequence 16, Appli

ALIGNMENTS

```
RESULT 1
US-08-152-922A-7
: Sequence 7, Application US/08152922A
: Patent No. 5395614
: GENERAL INFORMATION:
: APPLICANT: Knapp, Bernhard
: APPLICANT: Knapp, Erika
: APPLICANT: Enders, Burkhard
: APPLICANT: Kuemper, Hans
: TITLE OF INVENTION: Protective Plasmodium Falciparum
: TITLE OF INVENTION: Hybrid Proteins which Contain Part-Sequences of the Malaria
: TITLE OF INVENTION: Antigens HRP1 and SERP, the Preparation and Use Thereof
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
: ADDRESS: Dunner
: STREET: 1300 I Street, N.W. Suite 700
: CITY: Washington
: STATE: D.C.
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/152,922A
: FILING DATE: 16-NOV-1993
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/806,471
: FILING DATE: 13-DEC-1991
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Einaudi, Carol P.
: REGISTRATION NUMBER: 32,220
: REFERENCE/DOCKET NUMBER: 02481-1143-02000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-408-4400
: TELEFAX: 202-408-4400
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 115 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-152-922A-7
: Query Match 7.3%: Score 615; DB 1; Length 115;
```

Best Local Similarity 100.0%; Pred. No. 2.2e-26;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 KELTYPOLFDLTNMLTCDNHGKRYLDGVEEINELLYKNEFDLRLKNDVAND 199
Db 1 KELTYPOLFDLTNMLTCDNHGKRYLDGVEEINELLYKNEFDLRLKNDVAND 60

QY 200 YCOIPFNLIKIRANELDVLKLVFGYRKLPLNDIKDVGKMEYIKKKKTININE 254
Db 61 YCOIPFNLIKIRANELDVLKLVFGYRKLPLNDIKDVGKMEYIKKKKTININE 115

RESULT 2
US-08-353-700-1
Sequence 1, Application US/08353700
Patent No. 5599919
GENERAL INFORMATION:
APPLICANT: YEN, TIMOTHY J.
APPLICANT: RATTNER, JEROME B.
TITLE OF INVENTION: NUCLEIC ACID ENCODING A
TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: DANN, DOEFMAN, HERRELL AND SKILLMAN
STREET: 1601 MARKET STREET, SUITE 720
CITY: PHILADELPHIA
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,700
FILING DATE: 09-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: REED, JANET E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: HUMAN
US-08-353-700-1

Query Match 4.28; Score 356.5; DB 1; Length 3248;
Best Local Similarity 18.9%; Pred. No. 8.2e-11;
Matches 350; Conservative 304; Mismatches 631; Indels 569; Gaps 79;

QY 125 YADLKH- VANNYLLITKELKYPQDLFTNMLTCDNHGKFLIDG---YEELNELLY 179
Db 709 FSDQKHKEIENMLCKTSQTL-GQVEDL-EHKLQLSN-----EIMDRCYODLHA EYE 761

QY 180 KLVNFYDLRLKLVAND-----YCOIPFNLIKIRANELDVLKLVFGYRKLPLNDI 231
Db 762 SLR---DLKSKDASLYTNEHQSLLAFDQQRPMHHSFAN-----IIGEGSMFSE 810

QY 232 KDNVGMEDYIKKKKTININELIESKTTIDKNKNATKEEKKLYQAOYDLSYKQ 291

Db 811 RSECHLEADQSPKNSAILQNRVDSLFSLES-OKQMSDLOKCEELVQIGETIEENLMK 869
QY 292 LEEAH-NLISVLEKRIDPLKKN-----ENIKELDKINEIKNPPASGMPNLL 341
Db 870 AEOHQSFVAETSOIKSLQEDTSAHONVAVETLSALBNKKEILO-----LL 916

QY 342 DKNKIEEKEIKEIAKTIKFNIDS-----LFTDPLELEYLEKKNKNDISAKVETKE 396
Db 917 --NKKVEEQAEIQELKKSNNLSDLSLEKLOLSETLSLE--KREMSIISLN-KREEE 971

QY 397 STEPNETNGVTPPLSYDINDNALNELNSFDLNPEDYTPKSPKNITDNEKKKFINET 456
Db 972 LTOENG-----TLKEINASLNOEKM--NLIO-----KSESFANYID-EREKSISEL 1014

QY 457 KEKTIKIEK-----KKIESKRSYEDRS--KSLNITKEXELNLEIYDSKRNNDIDLNE 509
Db 1015 SDYKOEKLIILLOCEETGNAYEDLSQYKKAQENSKLECLNCSLCEKRNNELOL 1074

QY 510 EKMCKRYSYKVEKLTJHNTFASYENSK--HNLEKLTALK--YMEDYSLRNIYVEKELKY 566
Db 1075 KEAFKHEQEFUTKL---AFAERNQMLLELYQALSEMTDQNNKSEAGGLKQ 1130

QY 567 YKNLSKIENEIETLVENIKKDEQULFEKTKTQDE-----MKPEKILEYSDI-- 614
Db 1131 EIMTLKEBQNMOKREVDNLQENQOLMKVMKTKHECONLESEPIRNSYKERESEBQCNF 1190

QY 615 --YKVOYQKVLNKNKIDELKKTOLILKNVELK-----HNHV----- 649
Db 1191 KPOHDLVEKEIISDSYNAQVOLAEMLRNKLKLOESKEKECLOHELOTTRGDLEISNL 1250

QY 650 -----PNSYKQENKOEPPYLLVLKKEIDKLKVPKVEYL 684
Db 1251 QDMQSOEISGLKDEIDAEKTYISGPHELSTQNDNAHLQSLQTTNKLNE-LEKICEI 1309

QY 685 INEEKNKITGEGSDNSPSTBEGELTGATTKRQOAG-----SALEGBSVQ 731
Db 1310 LQAEKVELVETLINDSRSECTY-----ATRKMAEEVKKLNEVKIINDSGLLHGEI 1362

QY 732 ---AAQOEKQAOPPVPPVPEAKQVTPPAPVNNKTENVSKLDYEKLEFPLNTSYI 787
Db 1363 DIPGGEFGEQNEQHPVSL-----APLDESN-----YEHLLTIS-- 1386

QY 788 CHKYTLVSHSTMNKIL--KQYKTYKEESKLSLSCDPLDLLFNIONNIPVWSPFSL- 843
Db 1397 -DKVEQMHFAELQEKFLSLQSEHKILHDHQOMSS-----KMSLQIYVDSLK 1443

QY 844 -NNSLSQLFMETYEKEMVCNLYKLKNDKINKLLEBAKVSTYKVTSSSSMQ----- 895
Db 1444 AENLYLSTNLNFPQODLVKEMQGLIEGLVPLSSSCVDPSSSLSDSSFYRALLEQT 1503

QY 896 -PLSLTPQDKREVSAN-----DSTGSHNINNSLKLFINITSLG 933
Db 1504 GDMSLSLNLEGAVSANQCSVDEYFCSSLOIYVDSIKAKENLVSLNLRNFGQDLYKEMQLG 1563

QY 934 KKNKIYDEL-----IGOKSENEFEKILK---DSDTFYVN-ESFTNIVKSKADDI- 978
Db 1564 LEEGLVPLSSSCVDPSSSLSDSSFYRALLEQTGMSLSNLEGVASVANSQSVDEYF 1623

QY 979 -NSLNDSEKRRK-----LEEDINKLAKTLOLSPDLVYKYLKLERL 1018
Db 1624 CSSLOEENLTRKEPPASPAKAVELLESICEVYRSLKELEKMESSQGIKMKKEIOLEOL 1683

QY 1019 FD-----KKRTVGYKMQQIKKLTLLKQLESKLS-----LNKKHYNLON 1058
Db 1684 LSSERQELICLRKOYLSNEMQOQKLTSVTLKEMSKLAEEKQTEQLSLEVARLDQ 1743

QY 1059 FSV-----FENKKKAEIAETENTLENTKILKH----- 1087
Db 1744 LDLSRSLLIGIDTEDALQGRNESCDSK-EHTSETTERTPRHNDVHOICDQAOQDLNDI 1802

QY 1088 ---YKGLVK---YNGESSP-----LKTLSLSE----- 1107
Db 1803 EKITETGAVKPTGCSGSPQDTNVEPPGEDKTOGSSSECISELFSFGPNALVPMDFLQ 1862

QY 1108 ---ESITQEDVASENFKVLSKLEG---KLDNMLNEKKLSYSSGLH---HLIAEL 1157
 Db 1863 EDHNLQLRKETSINENRLIHVEDRDRKVESLLN---EMKELDSKTHLOEVQLMKRI 1918
 QY 1158 KEVINKRNYTGNPSNNTDNNALLESY-----1185
 Db 1919 EACIELEKIVGELKE--NSDISEKLEFSCDHQELQREVETSEGLNDSLEMHADKSSRED 1977
 QY 1186 -----KKFLPEGDVATVVSSESDPLEOSQPKRPSTHVGASNTITTSQ- 1231
 Db 1978 IGDVNAKVNDSMKRFLDVENESIRSEKASIEHE-----ALYLEADLEVVOETKEL 2029
 QY 1232 --NVDEVDVYIPIFGESPEDEYDILGOVYT-----GRAVPS-----VIDNLSKIE 1278
 Db 2030 CLEKDNENKOKIVICL---EEL---SVYTSERNOLRGELDMKKTALDOLSEKMK 2081
 QY 1279 -----NEYEVLYLKPLAGVYRSKLENNVMTFNVVKDILNSFRNFKENKYLE 1330
 Db 2082 EKTQELSHOSECHICIOVAEAEVKETEL---LOTLSDDVSELLKDKTHLOEKLOS-LE 2137
 QY 1331 SDLIPYDLTSSNVVVDPIFKLKEKR-----DKFLSSNYIKDSITDINFRAND 1381
 Db 2138 KD---SQALSLTKCELENOIAOLNKEKELLYKESSESLQARLSSESDYEKLNVSKALEALV 2194
 QY 1382 VLGYYKI-LS-----EKYKSDLS-----IKKYINDKOGENEKYLPLN 2140
 Db 2195 EKGFALRLSTOEVOVQOLRGITKLVRLIADKCOLHIAEKLEKRENDSDLKDEEN 2254
 QY 1421 IETLYKTVNDKIDLFVILHLEKAVLNTYKSNVE---VKIKEL-----NYLKT 1465
 Db 2255 LERELQSEENOEVLVIDAENSKAEVEFLTKQIEMARSLKIFELDLVTLNSENKMLTKQ 2314
 QY 1466 IODKLAPFKNNNVGADLSTDYNNHNLTKFLSTGWFENLAKTYLSNLDONLOGL 1525
 Db 2315 IOEQOGLSE-----DKLLSS-----FKSLLEKEQAEI 2344
 QY 1526 NISOHQCKKOPNSGCFRHLDERECKCLNLYKQEGDKCVENP-NPTCENN 1578
 Db 2345 QIKSESTAVEMQN-----QLEKELNVAVALCGDQELMKATEOSLDPRITEEH 2393

RESULT 3
 PCT-US95-16216-1
 : Sequence 1, Application PC/TUS9516216
 : GENERAL INFORMATION:
 : APPLICANT: Yen, Timothy J.
 : APPLICANT: Rattner, Jerome B.
 : TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
 : TITLE OF INVENTION: Expressed Kinetochore Protein, and Methods of Use
 : NUMBER OF SEQUENCES: 4
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Dann, Dorfman, Herrell and Skillman
 : STREET: 1601 Market Street Suite 720
 : CITY: Philadelphia
 : STATE: PA
 : COUNTRY: USA
 : ZIP: 19103-2307
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: PCT/US95/16216
 : CLASSIFICATION:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/353,700
 : FILING DATE: 09-DEC-1995
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Reed, Janet E.
 : REGISTRATION NUMBER: 36,252

TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (215) 563-4100
 : TELEFAX: (215) 563-4044
 : INFORMATION FOR SEQ ID NO: 1:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 3248 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: not relevant
 : TOPOLOGY: not relevant
 : MOLECULE TYPE: protein
 : HYPOTHETICAL: NO
 : ANTI-SENSE: NO
 : PCT-US95-16216-1

Query Match 4.28; Score 356.5; DB 5; Length 3248;
 Best Local Similarity 18.9%; Pred. No. 8.2e-11;
 Matches 350; Conservative 304; Mismatches 631; Indels 569; Gaps 79;

QY 125 YADLKH--VRNYLTITKELKYPQLFDLTNMLTLCDNIHGFYLLIDG---YEFINELLY 179
 Db 709 FSDQHQRQEIENMCLTKTSQLT-GOVEDL-EHKLDLSL---ETMDKRCYQDLHAEYE 761
 QY 180 KLNFFYDLIRAKLNDVCAND-----YCOIPFNLIKIRANELDVLLKLVGYRKPLDNI 231
 Db 762 SLR---DLKSKDASLYTNEDHQSLAFDQOPAMHHSFAN-----IIGEGSMSE 810
 QY 232 KDNVGMEDYIKKKKKTENINELIESKTDIDKNKATKEEKKLYQAOYDLSYKQ 291
 Db 811 RSEKRLQADQSPKSAIIONHVDLSLEFLES-OKOMSDLOKCEELVQIGETENIMK 869
 QY 292 LEEAH-NLISVLEKRIIDLKKN-----ENIKELDKINEIKNPPRANGMTPNLL 341
 Db 870 AEQHQSVAVETSORISKLOEDTSAHONVVAETLSALENKEKELQ-----LL 916
 QY 342 DKNKIEHEKEIKETIKTFINIDS-----LFTDPLELEYLREKKNIDISAKVETKE 396
 Db 917 --NKVETQEQAEIDKSKSNHLLDSLSKEQLSEITSLK-KKEMSIISLN-KREIEE 971
 QY 397 STEPERYNGVYTPLSINDINNALNELNSFGDLINPDYKPEPSKIYTTDERKKFTNEI 456
 Db 972 LTOENG-----TKEINASLNOEKM-NLIQ-----KSESFANYID-EREKSISEL 1014
 QY 457 KEKIRIEK---KKIESDKSYEDRS---KSLNITKYEKLNLMEIYDSKNNNIDLTNE 509
 Db 1015 SDQYKQEKLIILQCEETGNAYEDLSQYKKAQKNSKLBCLNCTSLCENKRNELPOL 1074
 QY 510 EKMGKRYSYVEKLTTHNTPFASYENSK--HNLEKLTALK-VMEDYSLRNIYVEKELKY 566
 Db 1075 KEAFKAEHQEFLTKL---AFAERNQNLMELETVOQALRSEMTDQNNKSKSEAGLQ 1130
 QY 567 YKNLSKIENEIEFLVENIKKDEBQLEFKKTKDE-----NKPDEKILEVSDI- 614
 Db 1131 EIMTLKEONKQKQEVNDLLOENQMLKVMKTKHECONLSEPIRNSYKRESEKCNF 1190
 QY 615 ---KVQYQVYLLNNKIDELKKTLQILIKNVKL-----HNHV----- 649
 Db 1191 KPQMDLEVKETISLDSYNAOLVQELAMLNKELKLOESEKEKECLOHETLOTIGOLETSNL 1250
 QY 650 -----PNSYKQENKQEPYLLVLKLEKIDKLVKMPKVESTL 684
 Db 1251 QDMQSOETISGLKDEIDAEKYYISGPHLSQSNDNMHLQCSLOTYNNKLNK-LEKICEI 1309
 QY 685 INEKKNIKTEGQSDNSEPTEGKITGOATTKPGQOAG-----SALEGDSVQ 731
 Db 1310 LQAEKVELVETLNDLSRSECT-----ATRKMAEEVGKLLNEVKIINDOSGLHGEIVE 1362
 QY 732 ---AQAOEQKQAPPVVPPVPEAKAQVPPRPAVNNKTEVNSLDVLEKYEFLNYSI 787
 Db 1363 DIPGEQEQNEQHPVSL-----APLDESNS-----YEHLTJLS- 1396
 QY 788 CHKYLVSHTMNEKITL---KQYKTKEEESKLSSCPLDLLFNIONNIPVYMSFDSL- 843

```

Db 1397 -DKVQHFAFLQKPLSLQSEHKILHDQHCMS-----KMSLQTYVDSLK 1443
QY 844 -NNSLSOLFMEIYKEMWCNLYKLDNDKIRKLLLEAKVSTSVKTLSSSSMQ----- 895
Db 1444 AENVLSTNLNLPQGDLYKQKQGLIEGLVPSLSSCVPDSSLSLSDSSFYRLLEOT 1503
QY 896 -PLSLTPQDKPEVAN-----DQTSHTNLNLSLKFENILSLG 933
Db 1504 GDMSLSLNLEGAVANOCSDVEFCSSLQTYVDSLKAEVLSTNLNLPQGDLYKQKQGL 1563
QY 934 KNNKIYQEL-----IQKSSNFYKILK---DSDFYV-ESTTNVKSADI- 978
Db 1564 LEEGLVPSLSSCVPDSSLSLSDSSFYRLLEOTGMSLLNLEGVVANOCSDVEFC 1623
QY 979 -NSLNDSKRRK-----LEEDINKLTKTLOSPDLKYKYLKLERL 1018
Db 1624 CSSLQDEULTKRTKTPAPAKVEELESCEVYRQSLKLEKMSQGLMKKKEIQELPOL 1683
QY 1019 FD-----KKTGYKQKQKILKLLKQLESKLNS-----LNPRHVLON 1058
Db 1684 LSEBROELDKRQKQYLSNEQMOCKLTSVTLKESKLAEEKQTEQLSLELVARLOLG 1743
QY 1059 FSV-----FPNKKKEAIAETENTLENTKILKH----- 1087
Db 1744 LDLSRSLGLIDTEDAIQGRNESCDISK-EHTSETTEPTPHDYOQDKDAQODLNDI 1802
QY 1088 ---YKGLVK---YNGRSP-----LKTLS- 1107
Db 1803 EKITETGAVKTCGSCGOSPDITVEPPGEDKTOGSSFCISELFSGPNALVPDLEQNO 1862
QY 1108 ---ESTQEDNVASLEPKVLSKLEG---KLKDNLEKKKLSYLSGLH---HLIAEL 1157
Db 1863 EDIHNLOLRVAKETSNENIRLHLVIEDRDKVESLNL---EMKELSKLHQEVOLMTKI 1918
QY 1158 KEVKKNNYNGNSPENNTPVNNALSEY----- 1165
Db 1919 EACTELERIVGELKKE-NSDLEKLEFSCDHQELLQREVTSEGLNSDLEMHAKSSRED 1977
QY 1186 -----KFLPECTDVAIVSESGSPDLEQSPKPPASTHYGAESNTITTSO- 1231
Db 1978 IGDVAVAVNDSMKRFLDVENELSRISERKASIEHE-----ALYIADLEVYQTEKL 2029
QY 1232 --NVDEVDVYIIVIFGESEEDYDLGQVYT-----GEAVTPS---VIDNLSKIE 1278
Db 2030 CLEKDNEKQKIVICL---EEL---SVYTSBRNOLRGELDMSKTTALDQLESEMK 2081
QY 1279 -----NEVEVYLKPLAGVYRSLSKQLENNVMTFVAVNKKDILNSFKRKNFKNVLE 1330
Db 2082 EKTQELSHQSECLHCIOVAEAEVKEKTEL---LQTLSSDVSELLKDKTHLOEKLOS-LE 2137
QY 1331 SDLIPYKDLTSSNVVVKDPYKFLNKKR-----DKFLSYNYTIKDSIDTDINFAND 1381
Db 2138 KD---SALSTITKCELENQALQNLKEKELVKESESLQARLSSESDYERKLVNSKALEALV 2194
QY 1382 VLGYYKI-LS-----EKYKSDLS-----IKKYINDKQGENEKLPLFNLN 1420
Db 2195 EKGEFALRLSTOEBVHQLRGIEKLRVIREADEKQOLHIEKLEKBERENDSKOVEN 2254
QY 1421 IETLYKVVNDKIDLFVHLEKLVNTYKESNV-----VKIKEL-----NYLKT 1465
Db 2255 LERELQSEENQOELVILDAEASAETLKTQIEEMARSLKIFELDLVTLRSEKENITKO 2314
QY 1466 IODKLADFKKNNNEVGIADLSTVYNNHNLTKPLSTGTGVFENILAKTVLSNLDNLOGL 1525
Db 2315 IQEGQGLSE-----LDKLLSS-----FKSLLEKEQDAEI 2344
QY 1526 NISOHQCVKQCPONSGCFRHLDERECCKLLNKGQGDKCVENP-NPTCENN 1578
Db 2345 QIKRESEKTAEMLCN-----QKELNEAVALACDQEIIMKATEQSUDPPIEEH 2393

```

```

; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DROITHE, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; EARLIER FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1786
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-8

```

```

Query Match          4.2%; Score 351; DB 4; Length 1786;
Best Local Similarity 20.0%; Pred. No. 7,7e-11;
Matches 269; Conservative 243; Mismatches 437; Indels 398; Gaps 68;

```

```

QY 229 DNIQDNGKMDYIKKKNTKTEINELIEES-KKTIQKKNKATPEEKKLYQAOYDLSTI 287
Db 757 ESVEENV--ESVAEENE--ESVAEENEESVAENVESVAPVEEIVASVESVAPSV 811
QY 288 YNKQLEE-----AHNLSVLEKRIQDTLKKENIKE-LLDKINETKNNPPANSGTPTNLI 341
Db 812 EESVAENVATNLSDNLSNLLGIEI-----EIKDSILNIEEYKE-----NVVTTIL 860
QY 342 DKKRIEHEKE-----IKEIKTIKFNIDSLFTDPLELEYLRKKNKIDISAKVETRES 397
Db 861 E--NVEETTESVTTTSNLEIEQEN--TITNTIEEK--LELHENV--LSAALENTQS 912
QY 398 TEPNEYNGVYPLSYNDINNALNELNS--FGDLINPDYTKERSKINUYDNERRKINE 455
Db 913 EEEKK-----EVIDVIEVEVEAVATLIEVEQAEEKSANT-----ITE 951
QY 456 IKEKIKIEKKIESDKRSYEDRSK-----SLNDIKKEYKILNELIYDSKPNNDITNEK 511
Db 952 IFE--NLEENAVESNENVAENLEKLNETVNTVLDKVEIYESGESLENEMDKAFSE 1009
QY 512 MGRKYSKVEKLTNHNTPFASYENSKHNLEKLTAKLYMEDYSLRNIVVERELKYKRNLI 571
Db 1010 I-----FDNVKIQIENL--LTGMPRSIE-----TSIVQSEKVDLN-- 1044
QY 572 SKINELIETLVENIKKQEOULEK--KITKDEKPPDEKILEVSDIVAVQV-----QKV 622
Db 1045 ---ENVVSSILDNIENKKEGLNKLLENISSTEGVQETVEHEQONVVDVAVPAMKQPL 1101
QY 623 LNMKIDELKKTQOLILKNVLEKNIHVPNSYKQEN-----KQEPYLLIVLKKETDKL 674
Db 1102 GILNEAGLKEKMFNLDV-----FKSISDVITYEIKEDVQKEVEKETYSII 1150
QY 675 KVPKRVESLINEKKNIKTEGOSDNSEPTGEGITQOATTKPQOAGSALLEGDSVOQA 734
Db 1151 EEMENETVDVLEEEKEDL--TDKMIDAVEESIE-----ISSDS--KEET 1190
QY 735 QEQKQAPPVAVPYPEKKAQVPTPAPVNNKTEVNSLDLYLEKLYFLNLSYICHTIIV 794
Db 1191 ESTDKERKDVSLVVEEVD-----NDMDESV-----EKVLELKN-- 1224
QY 795 SHSTNMEKILIKQYITTEESKSLSSCDPLDLFNIQNNIPVMSMPSLNSISOLFMEI 854
Db 1225 ---MEELMK-----DAVE-----INDITSKLEET 1247
QY 855 YE--KEWVCNLYKLDNDKIRKLLLEAKKVSTSVKTLSSSSMQPLSLTPQDKPEVSANDPT 913

```

```

Db 1248 OELNEVEADL--IKDMETKEL-----FKALSBS-----KEIIDAKDT 1285
Qy 914 -----SH--STLNNSLKLFENILSGKKNKIYOELIGKSS--ENFEYKILDSOTFY 963
Db 1286 LEKVEEHDITTLTLDVEVL-----KDVEEDKIEKVSOLKLEEDILKE----- 1330
Qy 964 NESFTNVKSKADINSLNDSKRRKLEEDINKLKTLOTLSFDLYNRYKLEKLEKFKKK 1023
Db 1331 -----VKRIKLESEFI--LEDYKEKLTETDLEKKE--EIKDKHPEKFEAEELKDLA 1302
Qy 1024 TVGKYYKMOIKRLTLKLEQESKLSLNPKHVLONFVFNKKKAEALTAETENTLENTKI 1083
Db 1383 DI-----LKEVSSLEVEEKKLEEVHELKE-----EVEHIIISGD-- 1416
Qy 1084 LKRYKGLVITYNGESSPLKTLSEESIOTEDNVASLENFK--VLKLEGLKNDLNEKK 1141
Db 1417 --AHIKGL-----EEDDLREVDLKGSLDMKLGDMELG--DMKE 1453
Qy 1142 KLSYLSGLHLILAELEKVIKKNKYTGNSPSENTDVNNALSEYKFLPEGTDAVATVSE 1201
Db 1454 SLEEVTKRLEGRVSLKDVLSA--LGMDEQMTKRKAQPKLEEVL-----LKEEYKE 1506
Qy 1202 SGGDTLEQS-----QPKKPASTHVGAESENTITTSQNVDEVDVITVPIFGESEEDYD- 1254
Db 1507 EPKKKIKKKKVFEDIKDKEPEDEIVEYEMKDEDIEEDVEEDIEEDIEDKVEDIDEDIDE 1566
Qy 1255 DLQGVYGEAVTSPVINDILSKINEVEVLYKLPLAGYRSKLOENNVTFNVANKDI 1314
Db 1567 DICE-----DKDEVIDILYOK--EKRIEYKAKK-----KKLEKREVEGVSGLKHVDEV 1614
Qy 1315 LN--SPENKR--ENFKNVLESDLPYKDLTSSNVVVDPPYKFLNKRKDRKDLSSYN--YIK 1369
Db 1615 MKYQKIDKEVDKVSALSKS-----NDVTN--VLKONQDFSKVK--NFVKKYYKFAA 1665
Qy 1370 DSIDTINFANDVLYGY-----KILSEKYSLDSIKYIKNDKOGENEKYLPELNN 1420
Db 1666 PETSAAVAFAFAYVGFETFLSFSSCVTIASSTYL--LSKYVDKFIN-----KNKERPFYSF 1718
Qy 1421 IETLYKTVNKKIDLFVYHLEKVLNITYEKSNSVEKIKELNLYKTIDOKLDFKNNNFV 1480
Db 1719 VFDFIKMLK-----HYLOOMKEKFSK--EKNNNVI 1746
Qy 1481 GIADLSTDYNNHNLTKFLSTGVFEN 1507
Db 1747 EYTKAKKKGKGVQYTKTEKTKYVDKN 1773

RESULT 5
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9901
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-328-254-6

Query Match 4.28; Score 350.5; DB 1; Length 2482;
Best Local Similarity 19.84; Pred. No. 1.2e-10;
Matches 355; Conservative 281; Mismatches 617; Indels 537; Gaps 79;

Qy 125 YADLKHR--VENVYLLTIKELKYPOLFDTNMLTLCNIGHFKYLLDG---YEEINELLY 179
Db 77 FSDQKHQKEINMGLKTSQLT--GOVEDL--EHKLQLSN-----EIMDKRCYODLHAYE 129
Qy 180 KLNFFDILRAKLNDVANCAND-----YCOIPNLKIRANELDVLKLYFGYKRPDLNI 231
Db 130 SLR--DLKSKDASLVYINEDHQSLLAFDOQAPMHHSFAN-----LIGEGSMPSH 178
Qy 232 KDNNYKMDYIKKKKKTENENNELIEESKRTIDKKNKATKEEKKKLYQAOYDLSTNNQ 291
Db 179 RSECRLEADQSPKSAIILQNVDSLEFSLES--OKOMSDLOKQCEBLYQIKGEIEBMLK 237
Qy 292 LEBAH--NLISVLEKRIIDLKKN-----ENIKELDKINEKNPPRANGTPTTL 341
Db 238 AEOHNSFVAETSORISKLOEDTSAHQNVVAETLSALENKEKEIQ-----LL 284
Qy 342 DKNKIEHEKEIEIKTIKFNIDS-----LFTDPLELEYLAEEKKNIDISKAVETKE 396
Db 285 --NDKVETEQAIEQLKSNHLSLSDSLKEIQLSETLSLE--KKEMSIIISLN--KREIEE 339
Qy 397 STEPNEYPGVTYPLSYDINNALNELNSFGDLINPFDYTRKPSKNITYONERKKFTINEI 456
Db 340 LTQENG-----TLKEINASLNQEM--NLIQ-----KSESFANTID--EREKSISEL 382
Qy 457 KEKIKIEK---KKIESDKSYEDRS--KSLNDITREYKILNEIYDSKFNNDIDLTNF 509
Db 383 SDQYKQEKLLILQRCSEETGNAYEDLSQKYKAQKNSKLECLLNECSTLCENRKNLEQL 442
Qy 510 EKMMGKRISYVEKLTJHNTPASTYENSK--HNLEKLTALK--YVEDYSLRNIYVEKELKY 566
Db 443 KEAFAKEHOEFITKL--AFAEERNONLMLELTVOALRSEMTDMONNSKSSEAGLKQ 498
Qy 567 YKNLSIKENIEITLVENIKKDEOLFEEKITTXDE-----NKPDEKILEVSDI-- 614
Db 499 EIMTLKEQNMKQKQEVNDLQENQILMKVMKTKHCONLSEPIRNSYKVERESERNOCNF 558
Qy 615 --KVQYQKVLNNKIDELKTLQILKNNELK-----HNHIV----- 649
Db 559 KPMQDLEKVELSDSYAQAQLVQLEAMLRNKELKLOESKEKECQOHLEQTLRGDLSTNL 618
Qy 650 -----PNSYKQENKOEPPYLLIVLKEKIDKLVKMPKYESTL 684
Db 619 QDMQSOEISGLKDEIDAEEKYISGRPHELSTQNDNHLQCSLOTATNNKLINE--LEKICEI 677
Qy 685 INEKKNNIKTEGQSDNSESPREGEITGOATTKPGQOAG-----SALEGSQVQ 731
Db 678 LQAEKVELVETLNDRSRCIT-----ATRKMAEVEGKLINEVKLLINDSGLLHAGELVE 730
Qy 732 ---AQAOEQKQAPVAVPVVPEAKAOPVTPPAVNNKNTENVSCLDYLEKLYEPLNTSYI 787

```

```

Db 731 DDPGEEGEOPNEQHPVSL-----APLDESN-----YEHLLTS-- 764
OY 788 CHKIIVSHSTMNKIL---KOYKTEKEESKLSS-----CDPLDLFNION 831
Db 765 -DKVEQMFALQEKFLSLQSEKILHDQHCQMSKSELOTIYDSLKAELVJSTLRN 823
OY 832 -----NIPVWMSFDSLNSLSQLFMEIYERKVCNLYKLKDKIKNL 875
Db 824 FQGLVKEMOGLBEGGLVPSLSSSCVDPSSSLSSLGSSFRAL---LEQGDMSLSLNL 880
OY 876 LEEKKVYSTVKTLSSSSMQPLSLTPDKPEVSA---NDPISHSTNLNLSL-KLFENILS 931
Db 881 EGAVSANQCSVDEYFCSLQSEENLRKRETPAPAKVEELSLCEVYRQSLKEEKMES 940
OY 932 LG--KKNITVOELLGOKSSENFYEKILKDSPTFYNESTFNVKKADINDNESKRRK 989
Db 941 QGIMKNEI--QELQOLLSSEQOELDKQYLSNEBW-----OQKLSVTLEMEKSL 992
OY 990 LEEDINKLKTLOLSPDLNKKLLEKLPDKKRTVGKQKQIKKTLTLKQLESKLNSL 1049
Db 993 AAEK-----KQEQSLSL-ELVARLOQGL-----DLSSRLSGIDTEDALQGR 1035
OY 1050 NNPKHVLONSVFPNKKKEAETENTLNTKILKH-----1087
Db 1036 N-----ESCDISK-EHTSETTEPKHDVHQICDKDAQODLNDIEKIT 1078
OY 1088 YKGLVK---YNGGESP-----LKTLSR-----E 1108
Db 1079 ETGLKLPGESESGSDPTNPEPGEDKTOGSSSECSISLSSGNALVPMFLGNOEDIH 1138
OY 1109 SIQTEDNVASLENKVLKLEGG---KLKDLNLEKKLSTYLSGLH---HLIAELKEVI 1161
Db 1139 NLQLRVKTSENMRLHLHVEDRORKEVSLN---EKKELDKLHLQEQVLMKRIACI 1194
OY 1162 KKNKTYGNSPBNNTDVNNALST-----1185
Db 1195 ELEKIVELKKE-NSDISEKLEFSCDHOELLQVETSEGLNSDLEMHADKSSREDIGD 1253
OY 1186 -----KKFLPEGDVATVSESGSDPTLEQSPKKPASTHVGESNTTTSO---NV 1233
Db 1254 VAKVNDSMKEFLVLENLSTIRSEKASIEHE-----ALYELADLEVYQTEKLEK 1305
OY 1234 DDEVDVYIIPFGESEEDYDDLOQVYT-----GEAVPSS---VIDNLSKIE--- 1278
Db 1306 DNEKKQKIYVCL---EEL---SVTSEKNOLRGELDTMSKKTALDQLEKMKERTQ 1357
OY 1279 ---NEVEVLYKRLAGVYRSLKQLENNVTFNVVNNKADILNSFPNKRKNVLESOLI 1334
Db 1358 ELESQSECLHCIOVAEAVEKTEL---LOTLSSDVSELLKDKTHQEKLOS--LEKD-- 1411
OY 1335 PYKDLTSSNVVNDPKYKFLNKKR-----DKFLSSVYIKDSIDPTDINFANDVLYG 1385
Db 1412 -SQALSITKCELENOJOLNKEKELVYKESLSQARLSSESQYKLNKALEALVKEGE 1470
OY 1386 YKI--LS-----EKYKSDLS-----IKKYINDKQENKYLPLNITEL 1424
Db 1471 FALSLSTOEVBHOLRGIEKLRVIRADEKQOLHIAEKLKERERENDSLDKQVLENLE 1530
OY 1425 YKTYNDITDLFVHLEKVLNITYEKSNE-----VKIKEL-----NYLKTQDK 1469
Db 1531 LQMSSENOELVILDAESKAQEVETLKTQIEEMARSLVFEGLDLTLSEKENLKOLOE 1590
OY 1470 LADPKRNNNFVGIADLSTDYNNHNLITKFLSTGWNFENLAKTVLSNLLDGLQGLMNSQ 1529
Db 1591 QGQSLF-----LOKLSS-----FKSLLEKEQALQIKE 1620
OY 1530 HQCVKQCPONGSGFRHLDEREKCKLLNKKQEGDKVEND--NPTCNENN 1578
Db 1621 ESKTAIVEMLN-----QKLEINFAVVAALCGDQELMKATQESLDPTBEH 1665

```

RESULT 6
US-09-104-324B-4

```

: Sequence 4, Application US/09104324B
: Patent No. 6232460
: GENERAL INFORMATION:
: APPLICANT: T feci, Ozlem; Sahin, Ugur; Pfrendenschuh, Michael
: TITLE OF INVENTION: Methods For Diagnosis And Treating Cancers,
: TITLE OF INVENTION: And Methods For Identifying Pathogenic Markers In A Sample
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSER: Fulbright & Jaworski LLP
: STREET: 666 Fifth Avenue
: CITY: New York City
: STATE: New York
: ZIP: 10103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
: COMPUTER: IBM
: OPERATING SYSTEM: PC-DOS
: SOFTWARE: Wordperfect
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/104, 324B
: FILING DATE: 25-June-1998
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/892, 702
: FILING DATE: 15-July-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Hanson, No. 6232460man D.
: REGISTRATION NUMBER: 30,946
: REFERENCE/DOCKET NUMBER: LUD 5491
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 318-3000
: TELEFAX: (212) 752-5958
: INFORMATION FOR SEQ. ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 976 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: US-09-104-324B-4

```

```

Query Match 3.98; Score 332; DB 4; Length 976;
Best local similarity 20.7%; Pred. No. 3.8e-10;
Matches 236; Conservative 204; Mismatches 354; Indels 344; Gaps 59;

OY 196 CANYCQIPF---NLKIRANELD---VLKRVFGYKRPDLNKKNNCKMEYIKKNNK-- 247
Db 41 CTEDLEPPFAKTLNKNNGENIDSDPALQKYNF---LP---VLEQVNSDCHVOEGLKDS 94
OY 248 TIEN-----INLEIESKKTIDKNNKATKE-----ERKKLYQAO---YDLSI 287
Db 95 DLENSEGLSRVFSKLYKEAEK-IKKMWSTPAELRQKESKQENKRTIIEORAKIOELOF 153
OY 288 YNK---QLEFANHLISVLEKRIIDTLKKNENIKELDKINEIKNPSPANGSPNPTLLDK 343
Db 154 GNEVYSKLKEEG-----IQENKDLIKENNATRHLCMLKE-----TCARSAEK 196
OY 344 NKTEHEHEKEIKELAKTIKFNIDSLF-----DPLELYYLREKKNK-----D 387
Db 197 TKTYEERETROYMDLNNNTEKMITAFOELRYQVQENSRLEMEHFKLDEYKIQHLEOE 256
OY 388 ISAVETKES-----TEPNEYPNGVYTPLSYNDINNALLNLSFGDLINFDYTKK 438
Db 257 YKKEINDKQVSLLOITKEKKNMDIFLL--ESRQKVNOL-----EKTKL 305
OY 439 PSKNITVDNERKKEFINIEKIKIEKKRIESDKRSYD---RSKSLNDITKREYKLLNE 494
Db 306 QSENLKOSIEKQHLTKLELEDIKVSLQSVSTOKALEEDLOIATKTIQQLTEERQOMEE 365
OY 495 IYDSKFNNNIDLTNFE-----KMGKRTSYKVEKLLHNHTFASYENSKRN--LEKTLKA 546
Db 366 SNKRAASHFVYTEFTTVCSLEELRLTEOORLEKNEODKILTMELQKSSLEEMTKL 425

```


QY 547 L-KYMEDYSLRNIVKEKELKYKLNLSKIEIETLIVENIKDEOLFEKKITKDNKP 604
 Db 426 TNKNEVELEKTKVGEKTELLYEN-----KQEKIAEELKGTGEQELIGLQAREKRVH 479
 QY 605 DEKLEVSQVAVOVAVLNNKIDELKKTOL-----ILKNVLKINIHVPNSYKQENKQEP 661
 Db 480 D--LET-OLVAITTSOQYVSKYKVDL-KTELENEKLTNLTSHC--NLSLSENKE-- 529
 QY 662 YLLIVLKEKIDKLKVPMPKVESLINEEKNKIKTEGOSDNSEPTGQATTGPGQA 721
 Db 530 -----LQOESDMTLELKNQOEDINNKKOEERMLKQIENLOETETOLRNE----- 575
 QY 722 GSALEGSVOAOEOQOAPVPEVPEAKAQVTPPAPVNNKTENV-SKLDTLEKLYE 780
 Db 576 -----LEVREBELKQKR-----DEVCKIKDKSEENNNLRKQYENKNTYIEELQ 620
 QY 781 FLNITSYCHKIYLVSHSTNMEKILKQYITKEESKLSGCDPLDLENNIPIVMSMF 840
 Db 621 -----ENKALKKKGTAEKQOLNVEI--KYNKLE--LEL----- 650
 QY 841 DSLNLSLOLMEI---YEKEMVCNLVKLKNDKI--KNILEEAKVSTSVKTLSSSSMQ 895
 Db 651 -----ESAKQKFGELTDTYOKET-----EDKKISEENLLEVEK--AKVIADKAV- 693
 QY 896 PLSLTPQDKPEVSANDTJSHSTNLSNLSKLEFENILSLGKNKNIYOELIGQKSENFYEKI 955
 Db 694 -----KLQKELDKRCQH-----KIAEWALMEKHQVQDKIEEDSE----- 731
 QY 956 LKQSDTYNESFTNFKVSKADINSLANDESKRRKLEEDINKLKTLOLSPDLNKKYKIKL 1015
 Db 732 -----LGLYKSKQEOESSL-----RASLEIELSLKAEI-----LSVKKOLEI 769
 QY 1016 ERLDPRKKTGVKQMOJKLKL-LKEEOLSKLNLNPKHVLQNPVFPNKKKAEIAT 1074
 Db 770 ER--EKE--KLREKENTATLKEKKDK-----TQTF----- 799
 QY 1075 ENTLNTRKILKHYKGLVKKYNGESSPLKTLSEESIOTEDNYASLENFVKLSLEGKLD 1134
 Db 800 --LEFPEIYWK-----LDSKAVPSQYISR-----NFTSVDH----- 829
 QY 1135 NLNLEKKLSTL-SSGHLHLIAELKEVAKNKNTGNSPS-----ENNTDVNNALESYK-- 1187
 Db 830 --GISKDKRYLMTSAKNTLSTPLP--KAYTVKTPTKRKLQORELNTPIEESKKR 882
 QY 1188 -----FLPECTQVATVSSSSGDTLEQ-SQPKKPATVHGAES-----NTITT 1229
 Db 883 KMAPEPINDSSSETDLSMWSE--BETIKTLYRNNNPASHLCVTKPKKAPSSLTP 938

RESULT 7
 US-08-687-080-51
 Sequence 51, Application US/08687080
 Patent No. 5965427
 GENERAL INFORMATION:
 APPLICANT: Gregory Dolganov
 TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
 NUMBER OF SEQUENCES: 175
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Dehlinger & Associates
 STREET: 350 Cambridge Avenue, Suite 250
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94306
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/687,080
 FILING DATE: 17-JUL-1996
 CLASSIFICATION: 435

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/592,126
 FILING DATE: 26-JAN-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Shultz, Charles K.
 REGISTRATION NUMBER: 38,615
 REFERENCE/DOCKET NUMBER: 4600-0111.30
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 324-0880
 TELEFAX: (415) 324-0960
 INFORMATION FOR SEQ ID NO: 51:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1312 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 INDIVIDUAL ISOLATE: TRANS. OF RAD50 CDNA (SEQ. 54), NT.
 INDIVIDUAL ISOLATE: 389 TO 4324
 US-08-687-080-51

Query Match 3.7%; Score 308; DB 2; Length 1312;
 Best Local Similarity 19.4%; Pred. No. 1e-08;
 Matches 248; Conservative 214; Mismatches 453; Indels 364; Gaps 56;

QY 411 LSYNDNNALNELSGDGL--NPFDTKPEKSNIIYDNERKKFINKIEKIRIEKKI 467
 Db 84 LQFRDYN-----GELIIVQSMVCTQSKKTEFTLEGVITRTFHGKRVSLSSCA 134
 QY 468 ESKKSYSDRSK--LNDITKETE-----KLNIETYSKFNNDITLTFEKMG 514
 Db 135 EIDREMISSLGVSAVLNNVIFCHQOESNMPLSQKALKQKFEDEFS-----A 182
 QY 515 KRYSKYVEKLTHTNMFASYSKSNHLEKLTAKLYMEDY-----SLRIVYEKE--LKY 567
 Db 183 TRYKALETLQVROTQOGKVEYOM-----LKYKQYKKEACEIRQDITSKEQLTSS 237
 QY 568 KNLISKIENEIETLIVENIKKDEOLFEKKITKDNKKPEKILEVSDIVKVOVK-----V 622
 Db 238 KEIYKSYENELDPLKNRKLEIHN--SKIMKIDNE-----IKALDSKKQMEKDNSELE 290
 QY 623 LMKKIDELKKTQLLKNVELKHNHVPNSYKQENKQEPYILVKKIEIDKLVMPKVE 682
 Db 291 EKMEKVFQGTDEOL-----NDLYHN-HQRTVREKERK-----LVQCHRELKLN--KES 336
 QY 683 SLINDEKKNITEG-----QSDNSE-----PSTGEITGQ 712
 Db 337 RLNQKESSELLVEQRLQLOADRHOEHIRARDLSLOSATOLELDGFERGPFSSQIKNF 396
 QY 713 ATRKPGQAGSA-----LEGDSVOAOEOQOAPVPEVPEAKAQVTPPAPVNNKTE 767
 Db 397 HKLVREOEGAKTANQLMNPAKETIKQO-----IDETRIDKTTGGRIETELKSE 448
 QY 768 NVSKLDVLEKLYEFLNITSYICHKIYLVSHSTNMEKILKQYITKEESKLSGCDPLDLF 827
 Db 449 ILSK-----KQNELKNV-----KYELQOLEGSSDRILELDELKAEKRELSKAEK----- 493
 QY 828 NIQNNIPVMSMPSLNSLSQLFMEIYERKEMVCNLVKLKNDK-----IKNLEPAKK 881
 Db 494 -----NSNVELKMEVTSIO-----NEKADLDRLTKLKDQOEQ 527
 QY 882 VSTSVKTLSSSMQPLSLTPQDKPEVSANDTJSHSTNLSNLSKLEFENILST-----GKNK 936
 Db 528 LNH--HTTTRQEMELTKDRKDKQDQIKIKSRHSDELTSIGLFPNKKQLEMDLHLSK 585
 QY 937 NIYO-----ELIGQKSENFYEKILKQSDTYNESFTNFKVSKADINSLANDESK 986
 Db 586 EINQTRDLAKLNKELASSQONKHNINNELKRKE-----EQLSSEYEDKLDFVCGSQDFESD 641

```

QY 987 RKLLEEDINKLK---TLQSFDLNKKYKRL-----ERLFDKKRYGKYMQI 1032
D 642 LDRLEKEIEKSSKRAMLAGATAVYSOFITQITDENOSCCVQCRVOTFELEQEVISDL 701
QY 1033 K-KITLKEQLESKLNINPKH-----VLQNSVFENKKKEAETAEENTLENTKI 1083
D 702 QSKRLAPDKLSTESSELKKEKRDEMLGLVPMQSI--DLKEKEIPELRNKLQNVNR 759
QY 1084 LKHYKGLVKYNGESSPLKLTSESIOTEDNVASLNFVKYLSKLEGLKKNLMEKKKL 1143
D 760 DIQKAKNDIE---BOETLGLTMBE---ESAKVCLDTVTMERQMLKO---VERKI 809
QY 1144 SYLSSGHLHLTAELKEVTKNNKNTGNSPSENNTDNNALSESRYKFLPBGTDVATVSESG 1203
D 810 AQQAAKLOGI--DLDRVQYQVNOEKOEHKLDIVYSSKIELNRKLI-----853
QY 1204 SDLEQSOFPKPASTHGAESNTTTS---QNVDEVDVYIYPIGESEEDYDLOGV 1259
D 854 QDOEOIOLHKTSTNELKSEKLIQISTNLORQOLEBOT---VELSTEVOSLYREIKD- 907
QY 1260 VTGEAVTPSVINDTLKSTENYEYVLYKLPLAGVYRSJKOLENNVMPFNVAWKDILNSRF 1319
D 908 -AKQVSP--LETLLEKHOE-----KEELINKKNTSNKTAQDLANDIK 948
QY 1320 NKRENFKNVLESDDLIPYKDLTSSNVV--KDPYK-----FLNK-----EKRDKFLS 1363
D 949 EKVNINGYM-----KDI--ENVIOGKDYKKOKETELNKYIAQSECEKKEKINE 999
QY 1364 SYNTIKSDIPD-----INFANDVL-----GYTKLS-----EK 1392
D 1000 DMRLMRDIDIDQKIOERWLODNLTLRKRNELKEVEERKOHLEKMGOMQVLOKMSEROK 1059
QY 1393 YKSDLSIKKIKINDKOGENEK-----LPFLNITETLYK-----TYND 1430
D 1060 LEENDIKRHHNHLALGROKYESEIIFKKELEBPQRPDEEYKREMIYMRTELNVK 1119
QY 1431 KIDLEVLHLEAKVLNVTYKSNVEKIKELNYLKTIODKLADFFKNNNF-----1479
D 1120 DLDIYVTLQDAIKMFH-----SMKMEIN--KIIRDLMRSTYRGDIEYIEIRSDAE 1171
QY 1480 -VGADLSTVDYNNHNLTK 1497
D 1172 NVSASDKRRNRYNVYMLK 1190

```

RESULT 8
US-08-592-126-148
Sequence 148, Application US/08592126
Patent No. 5821091
GENERAL INFORMATION:
APPLICANT: Gregory Dolganov
TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
NUMBER OF SEQUENCES: 151
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
City: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,126
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615

```

REFERENCE/DOCKET NUMBER: 4600-0111
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 148:
SEQUENCE CHARACTERISTICS:
LENGTH: 1312 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: Rad50, pro-translation of SEQ ID NO:54
US-08-592-126-148

Query Match 3.6%; Score 307; DB 2; Length 1312;
Best Local Similarity 19.4%; Pred. No. 1,2e-08;
Matches 248; Conservative 214; Mismatches 453; Indels 364; Gaps 56;

QY 411 LSYNDINNALNELSFGDLI--NPFDYTKPSKNITDNRKKFIMEIKIEKKKI 467
D 84 LQFRDVN-----GELLAVQRSNVCOTQSKTERFTLEGVITRRHGEKVSLSKCA 134
QY 468 ESDKRSYEDRSKS--LNDITKEYE-----KLNEIYDSKFPNNIDLTNFERKMG 514
D 135 EIDREMISSLGSAVLNNVIFCHQEDSNMPLSEKALKOKFDFIFS-----A 182
QY 515 KRYSYKVEKLTNHNTPASYSNSKHNLEKTKALKYMEDY-----SLRNIYERK--LKY 567
D 183 TRYIKALETLERQVROTQOGAKVEKOME---LKYLKQYKKEACEIRDQITSKEAQLTSS 237
QY 568 KNLISKIENELETLVENIKKDEBOLFEEKITIDENKDEKLESDLVKQVQVK-----V 622
D 238 KEIYKSTENELDPLKRNKKEIEHNL--SKTWKLDNE---IKALDSRKKQMEKDNSELE 290
QY 623 LLMNRIDELKQTOLILKNVELKHNHVPNSYKQENKQEPYVLYLKKREIDKLKVMRVE 682
D 291 EKMEKVGQGTDEQL-----NDLYHN--HQRTVREKERK-----LVQCHHELELN---KES 336
QY 683 SLINEKKNIKTEG-----QSDNSE-----PSTGEITGO 712
D 337 RLNQEKSELLVEQGRLOLQADRHQEHIRADSLIQSLATQLELDGERGPFSEROIKNF 396
QY 713 ATPKGOQAGSA-----LEGSVQAQAOEQKQAPVPVPPVPEAKQVPPPAVNNKTE 767
D 397 HKLYREROEGAKTANQANDFAKETLTKQO-----IDETRDKKTGIGRIIEKSE 448
QY 768 NVSKLDYLEKLYEFLNNTSYICHKYLIVSHSTNNEKILKQYKITEEESKLSGCCDPLDLF 827
D 449 ILSK-----KQNELKNV-----KLEIQLESSSDRIELEDOELKAKRELSKAK-----493
QY 828 NIQNNIPVMSMDPSLNSLSQLFMEIYEKEMVGNLYKLDKNDK-----IKNLEBAKK 881
D 494 -----NSNVELTKMEVYISLO-----NEKADLDRLLRKLDQEMEQ 527
QY 882 VSTSVKTLSSSSMQLSITPDQKPEVGSANDFTSHTNLNSLKFEENILS-----GKNK 936
D 528 LNH--HTTTRTQMELTKRADKDEQIRKISRSHSDDELTSILGFPKKOLEDMHLKSKS 585
QY 937 NIYO-----ELIGKSSSENFYEKILKSDTFFYNESFTNFKSKADINDSLNDESK 986
D 586 EINTORRLAKLNKELASSDEQKNHINNELKRR-----EQLSSTEDLFLVCGSDPESD 641
QY 987 RKLLEEDINKLK---TLQSFDLNKKYKRL-----ERLFDKKRYGKYMQI 1032
D 642 LDRLEKEIEKSSKRAMLAGATAVYSOFITQITDENOSCCVQCRVOTFELEQEVISDL 701
QY 1033 K-KITLKEQLESKLNINPKH-----VLQNSVFENKKKEAETAEENTLENTKI 1083
D 702 QSKRLAPDKLSTESSELKKEKRDEMLGLVPMQSI--DLKEKEIPELRNKLQNVNR 759
QY 1084 LKHYKGLVKYNGESSPLKLTSESIOTEDNVASLNFVKYLSKLEGLKKNLMEKKKL 1143

```

Db 760 DIOPLKNDIE---EQETLIGTIMPEE---ESAKVCLITVTIMERFQMEKLD---VERKRI 809
Oy 1144 SYLSSGLHLLAELEKVKNNKYTGNSPENNTOVNNALLESKKFLPPECTDVAIVVSESG 1203
Db 810 AQAQAKLGI--DLDRIVQOVNOEKOEHKIDITVSKTELNRKLI-----853
Oy 1204 SDLEBOSOPKRPASTHVAESNTITTS---QNVDEVDVITVPIFGESEEDYDLGOV 1259
Db 854 QDQOQIOHKLSTNELKSEKQIOISTNLOQRQLEOT---VELSTEVOSLYNEIND- 907
Oy 1260 VTGEAVTPSVIDNITLSKLENEYVLYKPLAGVRSKLKOLENNVMTFNVVVKDILNSRF 1319
Db 908 -AKEQVSP--LETTLEKFOQE-----KEELINKNTSNKIAQDKLNDIK 948
Oy 1320 NKREFRKAVLESIDLIPYDLTSSNVV--KDPYK-----FLNK-----EKDKPLS 1363
Db 949 EVKKNHGYM-----KDI--ENYIQGKDYKKELETKVIAOISECEKHEKINE 999
Oy 1364 SYNVIKSDIDTD-----INFANDYL-----GYKILS-----EK 1392
Db 1000 DMLRMRODIDPOKIOERVLQDNLTIRKNEBELKEVEERKQHLKEMGOMOVLOMKSEHOK 1059
Oy 1393 YKSDIDSTIKYIINDKGENEY-----LPLNNIETLYK-----TYND 1430
Db 1060 LEENIDNIRKNNLALGROKGYEEERIHFKKELRPOFRDAEKEYREMMIVARTETLVNK 1119
Oy 1431 KIDLEVIHLEKLVNTYKESNVEKIKELNYLKTIOKLDLDFKNNNF-----1479
Db 1120 DLDIYKTLDAIMKFH-----SKMEIN--KIIRLMRSTYRGODIEYIEIRSDAE 1171
Oy 1480 -VGIAIDLSTDYNNHNLTK 1497
Db 1172 NVSASDKRRNNYRVMLK 1190

RESULT 9
US-08-290-919-3
Sequence 3, Application US/08290919
Patent No. 5720959
GENERAL INFORMATION:
APPLICANT: HOLDER, ANTHONY A.
APPLICANT: BLACKMAN, MICHAEL J.
APPLICANT: CHAPPEL, JONATHAN A.
TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DABRY & CUSHMAN, L.L.P.
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,919
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9203821.5
FILING DATE: 22-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/00367
FILING DATE: 22-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 53 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label= X
OTHER INFORMATION: /note= "X = M and N, or N"
US-08-290-919-3

Query Match 3.6%; Score 301; DB 1; Length 53;
Best Local Similarity 100.0%; Pred. No. 4.7e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1570 NNPCTENNNGGCDADAKCTEEDSGSNGKRTCECTKPDSPYLPFDGIFCSSN 1621
Db 2 NNPCTENNNGGCDADAKCTEEDSGSNGKRTCECTKPDSPYLPFDGIFCSSN 53

RESULT 10
US-08-480-604A-6
Sequence 6, Application US/08480604A
Patent No. 5736139
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHVE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:


```

1 STREET : 220 MONTGOMERY STREET, SUITE 2200
2 CITY : SAN FRANCISCO
3 STATE : CALIFORNIA
4 COUNTRY : USA
5 ZIP : 94104
6
7 COMPUTER READABLE FORM:
8 MEDIUM TYPE: Floppy disk
9 COMPUTER : IBM PC compatible
10 OPERATING SYSTEM: PC-DOS/MS-DOS
11 SOFTWARE : Patent Release #1.0, Version #1.30
12 CURRENT APPLICATION DATA:
13 APPLICATION NUMBER: US/08/405,496A
14 FILING DATE: 16-MAR-1995
15 CLASSIFICATION: 424
16 PRIOR APPLICATION DATA:
17 APPLICATION NUMBER: US 08/329,154
18 FILING DATE: 25-OCT-1994
19 PRIOR APPLICATION DATA:
20 APPLICATION NUMBER: US 08/161,907
21 FILING DATE: 02-DEC-1993
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: US 07/985,321
24 FILING DATE: 04-DEC-1992
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: US 07/429,791
27 FILING DATE: 31-OCT-1989
28 ATTORNEY/AGENT INFORMATION:
29 NAME: INGOLIA, DIANE E.
30 REGISTRATION NUMBER: 40,027
31 REFERENCE/DOCKET NUMBER: OPD-01308
32 TELECOMMUNICATION INFORMATION:
33 TELEPHONE: (415) 705-8410
34 TELEFAX: (415) 397-8338
35 INFORMATION FOR SEQ ID NO: 6:
36 SEQUENCE CHARACTERISTICS:
37 LENGTH: 2710 amino acids
38 TYPE: amino acid
39 TOPOLOGY: linear
40 MOLECULE TYPE: protein
41
42 US-08-405-496A-6

```



```

OY 1175 NTDVNNALES--YKFLPECTDVATVVSSESGDILEOSQPKKPASTHVAGESNTITTSQN 1232
Db 2585 -VNLSALEGLFPMKSKVGTGETAT-----KNNTLPT-----2616
OY 1233 VDDEVDDVIVPIGESEEDYDGLGVVYTGAVT--PSVIDNLSKIE-----NE 1280
Db 2617 --DQVSSILLIPPYK-----DIFHLFCGSKSTKPKNKNTSIALIHIHISNRNIHG 2669
OY 1281 YEVYLLPGLGVYSLKKGKLENNVNFENVNPKDL-----NSRPMKRE 1323
Db 2670 CDFLYLENQNDIAISNNNNNSYSIFTHNKNTENNLIJCDISLIPYVIGKCPNKKLNPT 2729
OY 1324 NFKNV--LESDDLIPYKDLTSSNY-----VKKDPYKLNKKRDKFLSSNYI--- 1368
Db 2730 CFDEVYVVKQEDVPSKITTDADKYNTFSKDKLIGNILKNAISINNDEND--NTYTYLILP 2786
OY 1369 ----KDSIDPIDINANDVNLGYKILSEKYSDDLISIKKYINDKOGENE--KY-----LP 1416
Db 2787 EKFEELIDTRKVLACTGCDNKYIIHMKIEKSTMDKIK--IDEKTKICKICKYDVPTTKVA 2844
OY 1417 FLNNIEFLYKTV-----NDKI--DLFY--IHLKAVLNY 1446
Db 2845 TCEIIDIIDSSVLEKHEHTVYHSTLSRMDKLIITYPNEKTHFENFPVNPMLDKVL-Y 2903
OY 1447 TYEKS--NVE-----VKIKELNYLKTIDKLADFKKNNVGVGIADLSTDYN 1490
Db 2904 NYNKPIHIEHLPGAITTDYIDTRKTIQ--YLIRIPPY--HKDIHF-----SLEFN 2952
OY 1491 HNNLLTFSLTGAWFENLAKTVLSNLDGMLQGLNISQ-----HOCYKCKCPNNSGCFR 1545
Db 2953 NSLSLTK-QONONIIYGVNAKI-----FIHINOYKEIHHG-----DETKYS 2993
OY 1546 HLDRECKCLNLYKQEGDKCVENPPTCNENNG 1579
Db 2994 H-----LFTYSK--KPLPMDDOICNVITG 3015

RESULT 13
PCT-US93-07261-11
; Sequence 11, Application PC/TUS9307261
; GENERAL INFORMATION: PFEPM3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
; TITLE OF INVENTION:
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John H. C. Blasdale
; STREET: One Giralda Farms
; CITY: Madison
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07940-1000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07261
; FILING DATE: 19930805
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/927,531
; FILING DATE: 07-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Blasdale, John H. C.
; REGISTRATION NUMBER: 31,895
; REFERENCE/DOCKET NUMBER: DX0288K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-822-7398
; TELEFAX: 201-822-7039
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1588 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: protein
PCT-US93-07261-11

Query Match 3.5%; Score 296; DB 5; Length 1588;
Best Local Similarity 20.3%; Pred. No. 5,7e-08;
Matches 308; Conservative 248; Mismatches 601; Indels 362; Gaps 69;

OY 243 KKKNTKIENTINE--LIESKTIIDKNNKATPEEKKKLYQAOYDLSTYNKOLEAHNLIS 300
Db 12 EKNKARNALKEKKLKQKQKNDQAKADLTRKESQD-----SSSEKSLKEGVGEA 62
OY 301 VLEK-RIDTLEK-----NENIKELDKINEIKNPPANGTPTNLLDNKKIEEHEKEIK 355
Db 63 LKEKENNETLKKKLELQNKKEEKNKIKDNNDLAKKGNQKDKKIVPKKPESEVEDLK 122
OY 356 EIAETIKFNISLFTDPLEEYILREKKN--IDISAKVETESTEPNEDYNGVYTPLSYN 414
Db 123 EMELKEKEFIQHLKDYEE--RREKRRNMLRSLRDKLREIEOLEKLVNA----- 170
OY 415 DINNALNELSFGDLINPFDYTKPESKNIYTDNREKKFINKIKIEKKIESDKSY 474
Db 171 QLESAINELKERRASRRPMYVGMORGMKDEYDEWIKKYIDQAKKNGTKDEIKDKGDY 230
OY 475 ED-----RSKSLNDITKEYEKLEINEIYDSKFNNNIDLTNFEKMKGRYSYKVELT 525
Db 231 EIEYETKFGYGRNALGEL-DEYEE-----REK--KRYLAK-----264
OY 526 HNNTFASYENSKHNLKLTALK-----YMEDYSLRNIVEKELKYKNLSKIENEIET 580
Db 265 -----EDGEGLDKVEEKLFEETGYGFREKFPYTRILYARK--RNKQK-----305
OY 581 LVENIKQDEDLFEKKKITDKENKPKDEKLEVS-----IYVVOQYVLYNMKIDELKKTQ 635
Db 306 --KLDEK--EKKLAAEPEDEKKIKLKDSDDKVVAVN-----341
OY 636 LILKNVELKHNHVPNSYKQENKOEPPYL-----IVLKEIKDLKVPKVESLINEE 688
Db 342 -----KNKSPFDKFRAPDKRTMFRYRLSELFPYPRND-NELAVCGGSMOSKYNKG 392
OY 689 KKNITKQSDNSEPSTGEITGOATTKPGQOAGSALBGDSVQ-----AQA 734
Db 393 KL-----KSTENPFKRRRNKKLKERMQELHFKKYNKKYOKLE 431
OY 735 QEOKQAPVPVPEAKQVTPPAPVNNKTENV-----KLDEKLYEFLNTSYCH 789
Db 432 REKRENDGEPLNTPETHV--IRPSDLMQGENKSGHPKTYPTGKLEY-ESHVSK 487
OY 790 KYILVSHSTNNE-----KILQYKITKEESKLSQDPLDLFLFNIONNIPVMSHDS 842
Db 488 DYOL-EHEPPTKLPREYKGGHVSREYQDHPPTKLPEYE-----KGHVSREYQDNE 538
OY 843 LNSLSQLEMEIYKEMVCNLYKLNDKIKNLLKAKVSTSVKTLSSSSMOPLSLTPQ 902
Db 539 VRDLPE-----YEKGHVSREYQD-DNEGPESTLKEYDQTELAKKIDITNKPHE--SVDEX 590
OY 903 DKPEVSANDQSHSTNUNLSKLLENISLQKNKI-----QOL-----IGQ 945
Db 591 DQTELAKKIDITNKP--HESVDEYDQ--SELAKGKIDITNKPHEVSDEYDQTELAKGKETYN 647
OY 946 KSSSEFYEKILKSDQTFENESFTNFVSKADINSLNDKSRK-----KLEIDIKLKLT- 1000
Db 648 KPHENLEB--YNETDLAKGKREYTNKPHESVDEYDQ--SELAKGKIDITNKPHEVSDEYDQTE 704
OY 1001 LQISFDLYNKYKTLERLFDKKTGVKYMQIKKLTLLKEQLESL--NSLNNPKHYVLQ 1057
Db 705 LAKGEVYTNKARENMLEEYNETDLAKGK--EVTNKAARENMLEEYNETDLAKGKREYTNKAH--E 761
OY 1058 NFSVFPMKKKAELAEENTENTENKILKHKGLVKKYNGSSPLKLSIETEDNYA 1117
Db 762 NLEEY---NETDLA-----KGREYTNKAHENLEEYNETDLAKGKREYTNKA-----HE 805
OY 1118 SLNFKVLSKLEGLKLDNL--NLEKKKLSYLSGHLHLIAELKEVIKNNKNTGNSPSN 1174

```



```

Db      806 NLEEVNETDLAKGKEVTNKARENLEEVNETDLAKG-----KEVT-----NKAREN 850
Qy      1175 NTDVNNALLESKFF-LRPGTIVATVVSFGSDTLEQSPKRPASTHVAAESNTITTSQNV 1233
Db      851 -----LEEYNETDLAKGKEVTNKARENLEEVNETDLAKGKEVTNKAREN--LEEYNE 900
Qy      1234 DDEVDVIVIFEESEEDYDGLQVTVGEAVTPSVIDNIIKSTENEX---EVLTKPLA 1230
Db      901 TDLAKGKEVTNKARENLEEVNET-DLAKGKEVTNKARENLEEVKEEYKMKNNELQNKCD 959
Qy      1291 GYVRSK--KOLEN-----NVMTFNVVKDILNSRFNKRNFKN-----VL 1329
Db      960 GLKENAELKNELRNKSGDLKENAELKNELRNKSGDLKENAELKNELRNKSGDLKE 1019
Qy      1330 ESDLIPIKDL-TSSNYVVKDYPKFLNKEKRFSS-----YNTKISIDIDI 1376
Db      1020 ENAELKNELRNKSGDLKENAELKNELRNKSGDLKENAELKNELRNKSGDLKE 1079
Qy      1377 NFANDVL---GYKILSEKYSDDLSTIKKYINDROGENEYKLPFLNITETLYKTIVND-KI 1432
Db      1080 ELAKKELOKNG-----SEGLKENAEOKKELOKNGSEGLKENAELKNELRNKSGDLKE 1134
Qy      1433 DLEVIHLEAKYLYTYEKSNEYKIKEL-----NYLKT--IQDKLADFK- 1474
Db      1135 NAEIKNELRNKSGDLKENAELKNELRNKSGDLKENYVYTNNDLKNNDLQNKDLSNKD 1194
Qy      1475 -KNNFPGIADLSTIDYNNHNLTLFTLSTGWFENLAKTVLSN--LLDGNLOGMLNISHO 1531
Db      1195 MKNNELLNKDISNMDMKNNELKNLKNLDSN---EDMKNNELLNKDISNMDQON 1250
Qy      1532 CVKRCOP---QNSGCFRHLDERECKCLLYKQEGDKCVENPN-PTCENNNNGCDDADAK 1586
Db      1251 TGLKNTPSKQOQNTGLKNTPREROQNTGLKNTPSGQOQNTGLKNTPSGQOQNTGLKNTPN 1310
Qy      1587 CTEEDSG-----SNGKIT 1600
Db      1311 EROQNTGLKNTPSGQOQNT 1329

RESULT 14
PCT-US93-07261-16
; Sequence 16, Application PC/TUS9307261
; GENERAL INFORMATION:
; TITLE OF INVENTION: PFEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John H. C. Blasdale
; STREET: One Giralda Farms
; CITY: Madison
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07940-1000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07261
; FILING DATE: 19930805
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/927,531
; FILING DATE: 07-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Blasdale, John H. C.
; REGISTRATION NUMBER: 31,895
; REFERENCE/DOCKET NUMBER: DX0288K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-822-7398
; TELEFAX: 201-822-7039
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 1663 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; STRAIN: Malaysian Camp
; PCT-US93-07261-16

Query Match 3.5%; Score 296; DB 5; Length 1663;
Best Local Similarity 20.3%; Pred. No. 6e-08;
Matches 308; Conservative 248; Mismatches 601; Indels 362; Gaps 69;

Qy      243 KNNKTJENINE--LIESKTTIDKNNATYEEKKKLYQAYDLSYNKOLEAHNLIS 300
Db      12 EKNEKARNALKEKKIKDKQKNDQAKADLTFRKESOD-----SSSEKSLKERVNGEA 62
Qy      301 VLEK-RIDTLK--NENIKELDKINEIKNPPANGNTPTNLLDKNKKIEEHEKEIK 355
Db      63 LKEKENETLKKLEENCKEKEKKNKIKDNNDALKNKGNDKDKKIVPKRPESVEKDLK 122
Qy      356 EIAATYIFNIDSLFTDPLEEYVLEKKNK-IDISAVETKESTEPNEYPNGVYTPLSYN 414
Db      123 EMELKEKEFIQHLKDYEE---RKEKRRNWLRSRLRDKLREIOLKLA----- 170
Qy      415 DINNALNELNFGDLINPDYTKRPNKIYTDNKRKKFINKIKIKIEKKIESDKSY 474
Db      171 QLESAINELKERRASRRRMVYKQGMKDEYDEWIKRYDQAKNGCTKDEIKDKDGY 230
Qy      475 ED-----RSKSLNDITKEYEKLNEIYDSKFNNDIDLTNEKMKGRYSYVERLT 525
Db      231 EEIYETKFGYKRENAIGEL-DEYEE-----REK--KRYLYK----- 264
Qy      526 HHNTFASYENSKHNLKLTALK-----YMEDYSLRNIVKELKYKNLISKIENEIET 580
Db      265 -----EDGEGLKDYEEKLEETGYGFRKFPPTTLVYRK-----RNKEOK----- 305
Qy      581 LVENIKKDEBQLEFKKITIDENKPRDEKILKESD-----IYVYQYQVLYNNKIDELKKTQ 635
Db      306 ---KLKEDK---EKKLIAEPPDEKKIKLIKSDSDKVVPVFN----- 341
Qy      636 LILKNVELKHNHIVPNSYKQENKOEPPYL-----IVLKEIDKLVPMPKVESLINEE 688
Db      342 -----KNNSPFDPFRAPDKKRTMFYRLSELFPYVRKD-NELAVGDSMSKYNKG 392
Qy      689 KKNITKGQSDNSEPSTGEITGOATTKPGQAGSALGDSVQ-----AQA 734
Db      393 KL-----KSTFNPFKRRNRKIKERKMDELHFKKYYKRYQKLE 431
Qy      735 GEOKAQRPVVPVPEAKAQVPRPAPVNNKTENV-----KLDTLKEKLEFLNTSYICH 789
Db      432 REKRENDGEPLNTPELIH---IRPSDMDGKNSKGHPKTYOPTGLKEY-BESHVSK 487
Qy      790 KYILVSHSTNNE-----KILKQYITKEEESKLSSCDPLDLFNIONNIPVMSFMS 842
Db      488 DYOL-EHEPPTKLPREYKRGHVSREYQLDHEPPTKLPRE-----KGHVSREYQLDNE 538
Qy      843 LNNLSQLPMEITYEKEMWCNLKYLKNDKIKNLLEBAKKYSTVKTSSSSMOPLSITPQ 902
Db      539 VRDELPE---YEKGHVSRREYQL-DNEGSPSTLKEYDQTELAKGDKDITNKPHE--SVDEY 590
Qy      903 DKPEVSANDDTSHSTNNLSKLLENILSLGKNKI-----YQEL-----IGQ 945
Db      591 DQTELAKGDKITNKP--HESVDEYDQ--SELAGKADITNKPRESVDEYDQTELAKGKEVTN 647
Qy      946 KSENEFEKILKSDPTFYNESFTNPFVSKADDINSKADDSKAR---KLEEDINKLTKT- 1000
Db      648 KPHENLE--YNETDLAKGKEVTNKPRESVDEYDQ--SELAGKADITNKPRESVDEYDQTE 704
Qy      1001 LQLSFDLYNKYKLEKLEKLPDKKTVGKYKQIKKLTLLKQLEQLESKL--NSLNNPKHVLQ 1057
Db      705 LAKGKEVTNKARENLEEVNETDLAKG- EVTNKARENLEEVNETDLAKGKEVTNKAH--E 761

```

```

QY 1058 NFSVFNNKKKEAEIETENTLENTKILKHYKGLVKKYNGESSPLKLTSEESIOTEDNYA 1117
D 762 NLEEY-----NETDLA-----KGKEVNNKKAHENLEEVNETDLAKGEVTNKA-----HE 805
QY 1118 SLENFKVLSKIEGLKLNL---NLEKKLSTLSSGLHLHLAEKLVKNNKYNGNSPEN 1174
D 806 NLEEVNETDLAKGEVNNKKAHENLEEVNETDLAKG-----KEVT-----NKAREN 850
QY 1175 NTDVNNLESYKKE-LLEGTDVATVSESSDITLQSQPKKPAETHGAESENTITTSQNV 1233
D 851 -----LEEYNETDLAKGEVNNKKAHENLEEVNETDLAKGEVTNKAHEN---LEEYNE 900
QY 1234 DDEVDVYIPIFGESEEDYDDLQVYVGEAVTPSVIDNITLSIENY---EVLTKPLA 1290
D 901 TDLAKGEVNNKKAHENLEEVNET-DLAKGEVTNKAHENLEEVTEKDYMKNNELONKNSD 959
QY 1291 GYVRSK-----KOLEN-----NVMTFNVVNDILNSRFKNKFNK-----VL 1329
D 960 GLKENAEKLNKELNKGSDGLKENAEKLNKELNKGSDGLKENAEKLNKELNKGSEGLK 1019
QY 1330 ESDLIPIKDL-TSSNVYVKDYPKFLNKKRDKPLSS-----YNYIKDSIDTDI 1376
D 1020 ENAEKLNKELNKGSEGLKENAEKLNKELNKGSEGLKENAEKLNKELONKNSGGLKENA 1079
QY 1377 NFANDVL---GYRKILSEKYSDLSIKKYINDKGENEKYLPFLNNIETLYKTIVND-KI 1432
D 1080 ELKKELONKNG-----SEGLKENAEKLNKELONKNSGSEGLKENAEKLNKELNKGSDGLKE 1134
QY 1433 DLEVILHAEKLVNTYKRSNVEKIKEL-----NYLKT--IODKLADFK- 1474
D 1135 NAEKLNKELNKGSDGLKENAEKLNKELNKGSEGLKENYVNTNDLKNNDIONKDSNKD 1194
QY 1475 -KNNFPGINDLSTDYVHNNILTFELSTGMYFENLAKTVLSN--LLDGNLOGMLNISHQ 1531
D 1195 MKNNELLNKDISNMDKNNKELNKLDSN---EDMKKELLNKDIRNKDLSISNMEQN 1250
QY 1532 CVKROCP-----QNSGCFRHLDERECKCLILNYKQEDKCVENPN-PTCENENNGGCDADAK 1586
D 1251 TGLKNTSKOQONTGLKNTPREQONTGLKNTPSGOQONTGLKNTPSGOQONTGLKNTPN 1310
QY 1587 CTEEDSG-----SNGKKIT 1600
D 1311 EROQNTGLKNTPSGOQONT 1329

RESULT 15
US-08-446-855A-2
; Sequence 2, Application US/08446855A
; Patent No. 5849573
; GENERAL INFORMATION:
; APPLICANT: Stewart, Thomas S
; APPLICANT: Flores, Maria V
; APPLICANT: O'Sullivan, William J
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
; TITLE OF INVENTION: Phosphate synthetase II
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 1100 NO. 5849573th Giebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,855A
; FILING DATE: 06-Jul-1995
; CLASSIFICATION: 435

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Mitchard, Leonard C
; REGISTRATION NUMBER: 29,009
; REFERENCE/DOCKET NUMBER: 47-80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-446-855A-2

Query Match 3.5%; Score 293.5; DB 2; Length 2391;
Best Local Similarity 19.1%; Pred. No. 1.3e-07;
Matches 323; Conservative 252; Mismatches 543; Indels 577; Gaps 83;

QY 171 YEEIN-----ELVYKLNFPDILRAKLVNDVCANDYCOIPFNILKTRANLDVLKIVF 222
D 232 YKEINLPDPCNIDPLKYCNHFIRVI--KLNN-----ITYNK-----NKEEF 272
QY 223 GYRPLDINIKNVGMEDYIKNNKTIENINELIEESKKTIDKNNATKEEKKKLYQAQ 282
D 273 NY---TNMTINDSMEDHNEINGSISFNMC--PSISFDKS-----ESKNV----- 316
QY 283 YDLSTYKOLEEAHNLISVLEKRIDTLKKNENIKELDKINEIKNPPANSNGTPTNLLD 342
D 317 ---INHLLDKNNLITSSSEYIKDL-HNCNFSNSDK-----NDSFF 355
QY 343 KKKIEHEKIKELAKTIK---NIDSLTPDLELEYLLEKKNKNDISAKVETKESTE 399
D 356 KLYGCEYDKYLDLEENASPHYNNVD-----EYGYDVNKNNTNLSNKKIIONNNE 408
QY 400 PNEYPGVTVPLSYDINDNALLNEL-----NSFGDLINFPDYTEKPSKIYTD-NERK 450
D 409 NKK-----NKKNNNNNEVDYIKKDEDNVNNSKVPYSQYNNNANQNNHEHFNILNN 457
QY 451 KFINEIKERK-----IEKKIESDKSYEDRSKSLNDITKEYEKLNEYDSKFNNN 503
D 458 DYSTYIRKKMKNEEFLNVNKKRYDHRK-----IIVYDGGIKNS 498
QY 504 IDLTNFEKMGKRYSYKEKLTTHNTFASYE-----ASKHLEKLTAKLTMEYSLRNI 558
D 499 IIKNLIRHGMPLPLTYIIVPYYNFNHIDYDAVLISNCPGPKCDPLIKMLKDSLTENK 558
QY 559 VV-----EKEKLYYKNLISKIENETIETLEVENIKKDEO-----LPEKK 596
D 559 IIFGICLGNOLLGISLGDYTKMKRYGNR--GVNOPYQLVDNICYITSQHGICLKRS 615
QY 597 ITKDE-----NRPDEKILEVSD---IVKVOYQVYLLMKIDELKTKQTLILNVELKH 645
D 616 ILKRELALSYINNDKSIIEGISHKNGRFYSVQFHP---EGNNGPEBTSFLFNFL-- 669
QY 646 NIHPVNSTKOENKOEYVLIIVLKEI----- 671
D 670 --DIFNKKQYREYLGNIIYIKKKVILLGSGGLCIGAGEFDYSQVAILSKLECGIYV 727
QY 672 -----DKLVFMPK---VESLINEKKN--IKTEQSONSESESTE 708
D 728 ILVNPNIATVQTSGLADKV-YPLPVNCFEVEKTIKKRPPFLICTFG----- 774
QY 709 ITGQATTRPG-----QAGSALGDSVOAQ---AOEQQAOPVPVPE 750
D 775 --GGTALNCALMDQKVKLKKNNQOCLGTSLESIRIETNRLFLAKLEINERI-APYGS 831
QY 751 AKAOVPPPPADVNNKTENVSKLDY---LEKLYER--LNTSYI-----CHKYILVSH 796
D 832 AK-----NVNOAIDIANIKIGYPIIVRTTFSLGGLNSSPINNEEELIECKNKIFL-- 880

```



```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

```

```

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

```

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	194	3.9	500	AU088128	AU088128 AU088128
2	168.8	3.4	107	AU086832	AU086832 AU086832
3	158	3.2	500	AU087573	AU087573 AU087573
4	156.4	3.2	500	AU088129	AU088129 AU088129
5	119.8	2.4	365	N97742	N97742_11B3C3 czap
6	108.8	2.2	500	AU086246	AU086246 AU086246
7	101.6	2.1	400	N97689	N97689_1114C3 czap
8	98.2	2.0	354	N97605	N97605_1039C3 czap
9	88.4	1.8	313	T18122	T18122_0676C3 czap
10	82.8	1.7	282	T18003	T18003_0325C3 cbSE
11	78.8	1.6	1147	B13042	B13042_T30M24 -SP6
12	73.8	1.5	540	AZ813205	AZ813205_2M0080C2AP
13	71.6	1.4	1885	BEA220745	BEA220745_HMM002.B0
14	68.6	1.4	908	AZ548467	AZ548467_ENTEEX3J0TR
15	68.4	1.4	596	AZ640541	AZ640541_1M026512A4
16	68	1.4	907	CNS021J74	AL176953 Tetradodon
17	67.2	1.4	500	AU087665	AU087665 AU087665
18	66.4	1.3	827	CNS021S56	AL1176451 Tetradodon
19	65.8	1.3	731	AZ191502	AZ191502_SP_1020.A
20	65.6	1.3	960	CNS00K65	AL077453 Drosophilid
21	65.6	1.3	1106	BF264948	BF264948_HV_CEA001
22	65.4	1.3	783	CNS011I2	AL100400 Drosophilid
23	65.4	1.3	1135	CNS03GQ	AL226115 Tetradodon
24	65	1.3	885	AZ541657	AZ541657_ENTDV08TFR
25	64.8	1.3	641	AQ946120	AQ946120 Sheared D
26	64.8	1.3	1101	219	AL077453 Drosophilid
27	64.8	1.3	1282	80	BF264918 HV_CEA001
28	64.4	1.3	500	AU086975	BF264975 AU086975
29	64.4	1.3	843	AZ551618	AZ551618_ENTDV54TFR
30	64.2	1.3	1223	B12981	B12981_T24D11 -Sp6
31	63.2	1.3	885	AZ542280	AZ542280_ENTFEF69TFR
32	63	1.3	932	AZ689160	AZ689160_ENTIE54TFR
33	63	1.3	1277	146	BF264952 HV_CEA001
34	62.8	1.3	1147	80	BF264731 HV_CEA001
35	62.6	1.3	997	CNS0134P	AL102403 Drosophilid
36	62.6	1.3	1036	CNS03LMJ	AL250012 Tetradodon
37	62.4	1.3	394	AU060224	AU060224 AU060224
38	62.2	1.3	1101	219	AL063921 Drosophilid
39	61.8	1.3	848	AZ672705	AZ672705_ENTMB16TFR
40	61.8	1.3	870	AQ330286	AQ330286_obxb00465
41	61.6	1.2	671	AZ749598	AZ749598_RPCT-24-I
42	61.6	1.2	681	CNS02IOD	AL133990 Tetradodon
43	61.2	1.2	839	CNS012ZR	AL101037 Drosophilid
44	61	1.2	867	CNS0054A	AL057618 Drosophilid
45	61	1.2	871	AZ671126	AZ671126_ENTHPO5TFR

[illegible]


```

AU088129      500 bp      mRNA      EST      27-JAN-2001
LOCUS         AU088129 Sugano Malaria cDNA library Plasmodium falciparum cDNA
DEFINITION    clone XPFn6560 similar to P. falciparum (Nf) gene for merozoite
               surface antigen 1, mRNA sequence.
ACCESSION     AU088129
VERSION       AU088129.1 GI:12390270
KEYWORDS      malaria parasite P. falciparum.
SOURCE        Plasmodium falciparum
ORGANISM      Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE     1 (bases 1 to 500)
AUTHORS       Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.
TITLE         FULL-malaria: a database for a full-length enriched cDNA library
               from human malaria parasite, Plasmodium falciparum
JOURNAL       Nucleic Acids Res. 29 (1), 70-71 (2001)
MEDLINE       20574754
COMMENT       Contact: Junichi Watanabe
               Institute of Medical Science
               The University of Tokyo, Department of Parasitology
               4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
               Tel: 81-3-5449-5378
               Fax: 81-3-5449-5410
               Email: jwatanab@manage.ims.u-tokyo.ac.jp
               Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
               S. Construction and characterization of a full length-enriched and
               a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).
FEATURES
  source       1..500
               /organism="Plasmodium falciparum"
               /strain="3D7"
               /db_xref="taxon:5833"
               /clone="XPFn6560"
               /dev_stage="erythrocytic stage"
BASE COUNT    197 a      65 c      58 g      144 t      36 others
ORIGIN
Query Match   3.2%: Score 156.4; DB 107; Length 500;
Best Local Similarity 57.6%: Pred. No. 9.8e-29;
Matches 250; Conservative 0; Mismatches 184; Indels 0; Gaps 0;
Qy 3161 atccgaacacgctactgcagactctcagtgctctcaacaagaagaagccgaga 3220
    || || || || || || || || || || || || || || || || || || || ||
Db 57  AACCCATTAATGTTTACAAACTTTCTGTTNCTTAAACANTNMAAAGACGTGAAA 116
Qy 3221 tcgcgcgagacagagacactctggaacacacaaagatctctcaaacactacaaggcc 3280
    || || || || || || || || || || || || || || || || || || || ||
Db 117 TAGCAGAAACTGANAANNACATTAAAAAAACAGANTATTATTGAAACATTATAAGAC 176
Qy 3281 tcgcgaagttataatagcgagctctctctcgaagactctcgcgaggaagcctc 3340
    || || || || || || || || || || || || || || || || || || || ||
Db 177 TTGGTAATATTAATAATGTCGAATCATCTCAATTAATAAACTTTAAGTGAATCAATTC 236
Qy 3341 agaccggagataactacgcgcacgctcgaaactcaaggtctctgaagctcgaaagca 3400
    || || || || || || || || || || || || || || || || || || || ||
Db 237 AAACGAGAAGTAATTAATGCCAATTAAGANANNTAGAGTATTAAAGTNNATTAATGCAA 296
Qy 3401 agctgaagacaacactgaacttggaagaagaagctcgaactactctctcgaagcctc 3460
    || || || || || || || || || || || || || || || || || || || ||
Db 297 AACTCATATGATTAATTTACATTAGGAAGAAATAATATATCTTCTTAACAGTGGATTNC 356
Qy 3461 atcacccgctcgcgcgagctcaagaaagtcattgaagaacaagaactacaacgcaatgccc 3520
    || || || || || || || || || || || || || || || || || || || ||
Db 357 ATCTTTTAATTAATGATTAATAAGAGCTTATNNAANTTAANNTAATACAGCTNATTTCTC 416
Qy 3521 caacgcgaataataacagacgtgaataacgacacgagatcttaagaagaagttcctgctg 3580
    || || || || || || || || || || || || || || || || || || || ||
Db 417 CAAGTGAATTAATAAGAAAGTAAACAGACCTTNNAAATCTTAACNAAANNTTTTTCNCA 476
Qy 3581 aaggaacagatgtc 3594
    || || || || || || || || || || || || || || || || || || || ||

```

```

Db 477 NAAGCAAAAGTTNC 490
RESULT 5
LOCUS   N97742
DEFINITION    N97742 365 bp mRNA EST 18-NOV-1996
               118C3 czapFPD2.1, Debopam Chakrabarti Plasmodium falciparum cDNA
               clone PF1183C, mRNA sequence.
ACCESSION     N97742
VERSION       N97742.1 GI:1674760
KEYWORDS      malaria parasite P. falciparum.
SOURCE        Plasmodium falciparum
ORGANISM      Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE     1 (bases 1 to 365)
AUTHORS       Dame,J.B., Annot,D.E., Bourke,P., Chakrabarti,D., Christodoulou,Z.,
               Coppel,R., Cowman,A., Craig,A., Fischer,K., Foster,D., Goodman,N.,
               Hinterberg,K., Holder,A.A., Holt,D., Kemp,D., Lanzer,M., Lim,A.,
               Newbold,C., Ravetch,J.V., Reddy,G.R., Rubio,J., Schuster,S.M., Su
               X.-Z., Thompson,J.K., Vital,F., Wellens,T.E. and Werner,E.
               Current status of the Plasmodium falciparum genome project
JOURNAL       Mol. Biochem. Parasitol. 79, 1-12 (1996)
MEDLINE       97001675
COMMENT       Contact: Debopam Chakrabarti
               Department of Molecular Biology and Microbiology
               University of Central Florida
               Orlando, FL 32816-2360
               Tel: 407 384 2061
               Fax: 407 384 3095
               Email: dchak@pegasus.cc.ucf.edu
               Seq primer: T3.
FEATURES
  source       1..365
               /organism="Plasmodium falciparum"
               /strain="Dd2"
               /db_xref="taxon:5833"
               /clone="PF1183C"
               /clone.lib="czapFPD2.1, Debopam Chakrabarti"
               /lab_host="E. coli XL-1 blue"
               /note="Vector: Lambda ZAP II; Site_1: EcoR I; Site_2: Xho
               I; PolyA+ RNA, from asynchronous blood stage parasites of
               the Dd2 isolate cultured in vitro, was reverse transcribed
               using an oligo dt-Xho I primer. Second strand was
               prepared using RNase H and DNA polymerase I. EcoR I
               adapters were ligated to the cDNA, and it was digested
               with Xho I. Prepared fragments were ligated into EcoR I +
               Xho I digested lambda ZAP II vector."
BASE COUNT    163 a      39 c      43 g      120 t
ORIGIN
Query Match   2.4%: Score 119.8; DB 187; Length 365;
Best Local Similarity 60.7%: Pred. No. 1.9e-19;
Matches 196; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
Qy 3817 attgataacattcgtcccaaatcgagaagaatacgaagtcctctatcgaacctctg 3876
    || || || || || || || || || || || || || || || || || || || ||
Db 43  ATGATTAATATCTCTCTCAGATTGAAATAATGATGTTATATTAATAAACCCTTTA 102
Qy 3877 gcaagcgctcctaagctcctcaagaacaagctggagaaataacgctgaagcctcaatgtc 3936
    || || || || || || || || || || || || || || || || || || || ||
Db 103 GCTGAGATATTAAGAGCTTAAAAAACAAATTTGAAAAAACATTAATTATTAATTTA 162
Qy 3937 aacgtgaagacattcgaacgcgcgtcttaataagaagaanaattcaagaacgcttg 3996
    || || || || || || || || || || || || || || || || || || || ||
Db 163 AATTGAAACGATATCTTAATAATTCACGCTTTAAGAAACGAAATAATATTTCTTGATGATTA 222
Qy 3997 gagcgagactgattccctataaagacctgacccctcctaactacgtgtgtaagaccaca 4056
    || || || || || || || || || || || || || || || || || || || ||
Db 223 GAATCTGATTTAATGCAATTTTAACATATATCTCAATGATGATCATTTATGGAAGATTCA 282
Qy 4057 tacaagttcctcaataagaagaagaagataaattctcgtctgaattcaactatatacaag 4116
    || || || || || || || || || || || || || || || || || || || ||

```

Db	283	TTTAATATTATGATTCGAGAACAAAAACACACTTTTAAAGTTTACAAATATATTA	342
Oy	4117	gactccatcgacacccgatataca	4139
Db	343	GAATCAGTAGTAAATGATATTTAA	365
RESULT	6		
LOCUS	AU086246	500 bp	EST
DEFINITION	AU086246 Sugano Malaria cDNA library Plasmodium falciparum cDNA clone XPf02175 similar to P.falciparum gp190 (MSA, MSP1, PMMSA) for precursor of majo		
ACCESSION	AU086246		
VERSION	AU086246.1	GI:12388387	
KEYWORDS	EST.		
SOURCE	malaria parasite P. falciparum.		
ORGANISM	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.		
TITLE	FULL-malaria: a database for a full-length enriched cDNA library from human malaria parasite, Plasmodium falciparum		
JOURNAL	Nucleic Acids Res. 29 (1), 70-71 (2001)		
MEDLINE	20574754		
COMMENT	Contact: Junichi Watanabe Institute of Medical Science The University of Tokyo, Department of Parasitology 4-6-1,Shirokanedai, Minatoku, Tokyo 108-8639, Japan Tel.: 81-3-5449-5378 Fax: 81-3-5449-5410 Email: j.watanab@nagane.ims.u-tokyo.ac.jp Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Suganome,S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library Gene 200 (1-2), 145-156 (1997).		
FEATURES	Location/Qualifiers		
Source	1..500		
	/organism="Plasmodium falciparum"		
	/strain="3D7"		
	/db_xref="taxon:5833"		
	/clone="XPf02175"		
	/clone_lib="Sugano Malaria cDNA library"		
	/dev_stage="erythrocytic stage"		
BASE COUNT	244 a	69 c	72 g
ORIGIN	115 t		
Query Match	2.2%	Score 108.8	DB 107; Length 500;
Best Local Similarity	58.2%	Pred. NO. 1.3e-16;	
Matches 191;	Conservative	0;	Mismatches 137; Indels 0; Gaps 0;
Oy	1745	agagcgtctgttgagacac	1804
Db	19	ATACATATTAATTAAACG	78
Oy	1805	aagaagaaaataaaccag	1864
Db	79	AAGGACTACACACATT	138
Oy	1865	tgcgaagagtgctccctc	1924
Db	139	TACAAAAAGTTTATTA	198
Oy	1925	agaagctgagtgatlaa	1984
Db	199	AAAAAGCAACACTTAA	258
Oy	1985	aggaacatctactctcgc	2044
Db	259	CAGAACCATATTTATTA	318
Oy	2045	ccaagctgagagcctgat	

DB	319	CAAAGCTAAACACATGTTAAAGAAAGA	346
RESULT	7		
LOCUS	N97689	400 bp	mRNA
DEFINITION	1114C3 czapprepD2.1, Debopam Chakrabarti Plasmodium falciparum	EST.	18-Nov-1996
ACCESSION	N97689	clone PF114C, mRNA sequence.	
VERSION	N97689.1	GI:1674723	
KEYWORDS	EST.		
SOURCE		malaria parasite P. falciparum.	
ORGANISM		Plasmodium falciparum	
REFERENCE		Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
AUTHORS		1 (bases 1 to 400)	
		Dane, J. B., Anco, D. E., Bourke, P., Chakrabarti, D., Christodoulou, Z., Coppel, R. B., Comnan, A., Craig, A., Fischer, K., Foster, J., Goodman, N., Hiltnerberg, K., Holder, A. A., Holt, D., Kemp, D., Lanzer, M., Lim, A., Newbold, C., Ravetch, J. V., Reddy, G. R., Rubio, J., Schuster, S. M., Su, X.-Z., Thompson, J. K., Vital, F., Wellens, T. E. and Werner, E.	
TITLE		Current status of the plasmodium falciparum genome project	
JOURNAL		Mol. Biochem. Parasitol. 79, 1-12 (1996)	
MEDLINE		97001675	
COMMENT		Contact: Debopam Chakrabarti Department of Molecular Biology and Microbiology University of Central Florida Orlando, FL 32816-2360 Tel: 407 384 2061 Fax: 407 384 3095 Email: dchakepegasus.cc.ucf.edu Seq primer: T3.	
FEATURES			
source		Location/Qualifiers	
		1..400	
		/organism="Plasmodium falciparum"	
		/strain="Dd2"	
		/db_xref="taxon:5833"	
		/clone="PF114C"	
		/clone_lib="czapprepD2.1, Debopam Chakrabarti"	
		/lab_host="E. coli XL-1 blue"	
		/note="Vector: Lambda ZAP II; Site 1: EcoR I; Site 2: Xho I; PolyA+ RNA, from asynchronous blood stage parasites of the Dd2 isolate cultured in vitro, was reverse transcribed using an oligo dt-Xho I primer. Second strand was prepared using RNase H and DNA polymerase I. EcoR I adapters were ligated to the cDNA, and it was digested with Xho I. Prepared fragments were ligated into EcoR I + Xho I digested lambda ZAP II vector."	
BASE COUNT	177 a	58 c	53 g
ORIGIN		112 t	
Query Match	2.1%	Score 101.6;	DB 187; Length 400;
Best Local Similarity	60.5%	Pred. No. 8.3e-15;	
Matches 167; Conservative 0;	Mismatches 109;	Indels 0;	Gaps 0;
0y	3316	aagactctctccgaggaagcatccagccgagataactacgcgaagctcgagaactc	3375
Db	2	AAAACTTTAAGTGAAGTATCAATTTAAACAGAGTAATTAATTCACAAATTTGAAAAAATTT	61
0y	3376	aagcgcctctaaactgaagcaagctgaagcaaccggaagctcgagaagaag	3435
Db	62	AGAGATTTAAGTAATAATGATGGAATAACTCAATATTAATTTACATTTTGGAAGAAAAA	121
0y	3436	ctcagctaccctctagcggagctgcatcaactgatacgccgagctcgaagaagtcattaag	3495
Db	122	TTATCTTTCTTATACAAAGTGAATGATCATCATTTAATTTAACTGATTTAAAAAGAGTAATAAA	181
0y	3496	aacaagaactcacacgggaataagccccgaagcgagataatacaagcgtaataaagcactg	3555
Db	182	AATTAATAATTTATACAGGTAATTTCTCAAGTGAATAATTAAGAAAGTTAAAGGAAGCTTTA	241
0y	3556	gaatttacaagaagttcctgctcgtgaagaagacag	3591

DB	242	AAATCTAGCAAAATTTTCTCCAGACGAAAGTT	277
RESULT	8		
LOCUS	N97605	354 bp	EST
DEFINITION	1039C3 czapPPD2.1, Debopam Chakrabarti Plasmodium falciparum	CDNA	
ACCESSION	N97605	GI:1674623	
VERSION	N97605.1	GI:1674623	
KEYWORDS	EST.		
SOURCE	malaria parasite P. falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota, Alveolata: Apicomplexa: Haemosporida: Plasmodium. 1 (bases 1 to 354)		
AUTHORS	Dame, J.B., Annot, D.E., Bourke, P., Chakrabarti, D., Christodoulou, Z., Coppel, R., Cowman, A.E., Craig, A., Fischer, K., Foster, J., Goodman, N., Hildebrand, K., Holder, A.A., Holt, D., Kemp, D., Lanzer, M., Lim, A., Newbold, C., Ravetch, J.V., Reddy, G.R., Rubio, J., Schuster, S.M., Su, X.-Z., Thompson, J.K., Vital, F., Wellens, T.E. and Werner, E.		
TITLE	Current status of the Plasmodium falciparum genome project		
JOURNAL	Mol. Biochem. Parasitol. 79, 1-12 (1996)		
MEDLINE	97001675		
COMMENT	Contact: Debopam Chakrabarti Department of Molecular Biology and Microbiology University of Central Florida Orlando, FL 32816-2360 Tel: 407 384 2061 Fax: 407 384 3095 Email: dchakepegasus.cc.ucf.edu Seq primer: T3		
FEATURES			
Source	Location/Qualifiers 1..354 /organism="Plasmodium falciparum" /strain="Dd2" /db_xref="taxon:5833" /clone="PF1039C" /clone_lib="czapPPD2.1, Debopam Chakrabarti" /lab_host="E. coli XL-1 blue" /note="Vector: Lambda ZAP II; Site_1: EcoR I; Site_2: Xho I; PolyA+ RNA, from asynchronous blood stage parasites of the Dd2 isolate cultured in vitro, was reverse transcribed using an oligo dt-Xho I primer. Second strand was prepared using RNase H and DNA polymerase I. EcoR I adapters were ligated to the cDNA, and it was digested with Xho I. Prepared fragments were ligated into EcoR I + Xho I digested lambda ZAP II vector."		
BASE COUNT	151 a 50 c 49 g 103 t 1 others		
ORIGIN			
Query Match	2.0%; Score 98.2; DB 187; Length 354;		
Best Local Similarity	61.6%; Pred. No. 5.9e-14;		
Matches 157; Conservative 0; Mismatches 98; Indels 0; Gaps 0;			
Oy	3337 atccagaccgagataactaagcgcagcctcgagaacttaagtcctgcttaagtcgaa	3396	
Db	18 ATTCAAAACAGAAAGAAATTATGCCCAATTTAGAAAAATTAGAGTATTAAATAATAGAT	77	
Oy	3397 ggcgaagctgaagagaaacctgaacctgagagaagaagtcgaagtcctctagtcgga	3456	
Db	78 GGAAGACCAATGATTAATTTACATTATAGGAAAGAAAAATTATCTTTCTTATCAAGTGA	137	
Oy	3457 ctgcataccatgatacgccgagcgtcaagaagcattaaagaacaagaatacaccgcaat	3516	
Db	138 TTACATCATTTTAATACGCAATTAACAGAGTAATAAAAAATTAATATACAGTAAT	197	
Oy	3517 agcccaagcgaataataacagagcgtgataaataacgcaactcttaacaagaagtcctg	3576	
Db	198 TCTCAAGTGAAGAAATTAAGAAAGTCTTAAGAAATCTTACGAAAAATTTTCTC	257	
Oy	3577 cctgaagaaacagat	3591	

Db	258	CCAGAAAGCAAAAGT	272
RESULT	9		
LOCUS	T18122	313 bp	mrna
DEFINITION	0676c3 czappd2.1, Debopam Chakrabarti Plasmodium falciparum	EST	30-AUG-1994
ACCESSION	T18122		
VERSION	T18122.1	GI:462908	
KEYWORDS	EST.		
SOURCE	Malaria parasite P. falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE	1 (bases 1 to 313)		
AUTHORS	Chakrabarti, D., Reddy, G.R., Dame, J.B., Almitra, E.C., Lalpis, P.J., Ferl, R.J., Yang, T.P., Rowe, F.C., and Schuster, S.M.		
TITLE	Analysis of Expressed Sequence Tags from Plasmodium falciparum		
JOURNAL	Mol. Biochem. Parasitol. 66, 97-104 (1994)		
MEDLINE	95075403		
COMMENT	Contact: Debopam Chakrabarti Department of Molecular Biology and Microbiology University of Central Florida Orlando, FL 32816-2360 Tel: 407 384 2061 Fax: 407 384 3095 Email: dchak@pegasus.cc.ucf.edu Seq primer: T3.		
FEATURES			
SOURCE	1. 313 location/Qualifiers /organism="Plasmodium falciparum" /strain="Dd2" /db_xref="taxon:5833" /clone="0676c" /clone_id="czappd2.1, Debopam Chakrabarti" /lab_host="E. coli XL-1 blue" /note="Vector: Lambda ZAP II; Site.1: EcoR I; Site.2: Xho I; PolyA+ RNA, from asynchronous blood stage parasites of the Dd2 isolate cultured in vitro, was reverse transcribed using an oligo dt-Xho I primer. Second strand was prepared using RNase H and DNA polymerase I. EcoR I adapters were ligated to the cDNA, and it was digested with Xho I. Prepared fragments were ligated into EcoR I + Xho I digested lambda ZAP II vector."		
BASE COUNT	144 a 35 c 35 g 98 t 1 others		
ORIGIN			
Query Match	1.8%; Score 88.4; DB 188; Length 313;		
Best Local Similarity	57.1%; Pred. No. 1.8e-11;		
Matches 180; Conservative 0; Mismatches 133; Indels 3; Gaps 1			
OY	1574	agaacatgacacacattacacctgctgcatcctcatgagatcttaagcaatacttgaga	1633
Db	1	AAAAACAAGATATATATATATATATTTCTCTCTCTAATATATCTGATATATAGTTCAAA	60
OY	1634	agctcaccaaaagctcttaagtaatacgagaggaactctctcgcaaacattgttgaga	1693
Db	61	AATTAATAAAGGCTCTTTCATATCTTGAAGATATATCTTTAAGAAAGAATTTCTGAAA	120
OY	1694	aagaactaagatattacaagaatctcataagtaagatcgaaaaagagatcgagcgttg	1753
Db	121	AAGATTTAATCATATATATATATCTTGAAGAAAGCGCCTCGAGACCTGATATATAAAAAATTA	180
OY	1754	ctggaacattgaagaagaatgaagaacgctgtcttgagaagaagatltcaaaagacgaa	1813
Db	181	CAGAAGAATATAAGGCTAGTGTGAAAAACAATAATCTTGAAAAAAATTTTAAAGGACTTAAC	240
OY	1814	ataaacacgatgagagaagatcccgagaggtctccgatattgttaagctccagtcgaga	1873
Db	241	ATTGAGCAAAAG---CTTCCTTAGAAGATATCTGATATTTGTAATAATTAACAGTACAAAAG	297
OY	1874	tgctccctcattgaaaca	1888

[illegible]

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel.: 801 585 5606
Fax: 801 585 7177
Email: ddunne@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0080 row: C column: 24
Seq primer: CGTTGTAACGACGCGCCAGT
Class: plasmid ends
High quality sequence stop: 540.
Location/Qualifiers
I. .540

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C2M0080C24"
/clone_id="Mouse 10kb plasmid U06C1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (g11473114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

```

[illegible]

RESULT 13
 BE420745
 LOCUS
 DEFINITION HM002.B02 ITEC HM Barley Leaf Library Hordeum vulgare cDNA clone
 ACCESSION HM002.B02, mRNA sequence.
 VERSION BE420745
 KEYWORDS BE420745.1 GI:9418588
 SOURCE EST
 ORGANISM barley
 Hordeum vulgare
 Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticeae; Hordeum
 1 (bases 1 to 1885)
 REFERENCE
 AUTHORS Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier
 'S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,
 Hermann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P.,
 Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, T.,
 Pechioni, N., Quilset, C., Schuch, W., Selvaraj, G., Shariflou, M.,
 Sorrells, M., Warburton, M. and Wenzel, G.
 International Triticeae EST Cooperative (ITEC): Production of
 Expressed Sequence Tags for Species of the Triticeae
 Unpublished (2000)
 COMMENT
 JOURNAL Contact: Herrmann RG
 Botanisches Institut der LMU
 Menzinger Str. 67, D-80638 Munchen GERMANY
 Fax: 49 30 171683
 Email: hermann@botanik.biologie.uni-muenchen.de
 International Triticeae EST Cooperative (ITEC)
 http://wheat.pw.usda.gov/genome.
 Location/Qualifiers
 1..1885
 FEATURES
 source
 /organism="Hordeum vulgare"
 /cultivar="Barke"
 /db_xref="taxon:4513"
 /clone="HM002.B02"
 /clone_1id="ITEC HM Barley Leaf Library"
 /lssue_type="leaf"
 /dev_stage="14 day old"
 /note="Vector: pBluescriptSK(-); 850 bp average insert
 size."
 BASE COUNT 1138 a 219 c 212 g 176 t 140 others
 ORIGIN
 Query Match 1.48; Score 71.6; DB 167; Length 1885;
 Best Local Similarity 37.68; Pred. No. 6.4e-07;
 Matches 362; Conservative 0; Mismatches 602; Indels 0; Gaps 0;
 QY 631 aagatcgagcgcaagctgtgagctatgtgaagagctgtgcttcggtatcgcaagct 690
 |||
 Db 892 AA 951
 QY 691 ctgcacacatcaagcaatgtggaagaatggaatatataaagaataagaag 750
 |||
 Db 952 NNN 1011
 QY 751 accatcgagaacatgaagctgtatcgagaatccaaagaacatagacaaaataag 810
 |||
 Db 1012 AA 1071
 QY 811 aatgcaacgaaggaagaagaagaagctgtacccgagccgaactgacactgtcctc 870
 |||
 Db 1072 AA 1131
 QY 871 tataacaacagctgaagagccataactcatcagctgactggaagagcgcatagac 930
 |||
 Db 1132 AA 1191
 QY 931 accctcaagaagaatgaatatatacaagaactgtctcgacaagatgaataaagaat 990
 |||

Db 1192 AA 1251
 QY 991 cctccgcagccgaactctcggaacacccctaacagcgtcgtggaagaagaagata 1050
 |||
 Db 1252 AA 1311
 QY 1051 gagagcaggaagaagatcaagaagatcgccaacacataagttcaacatagattct 1110
 |||
 Db 1312 AA 1371
 QY 1111 ctcttactgacctcccttgagctgagctgactacttgaggaagaagaatagatgac 1170
 Db 1372 AA 1431
 QY 1171 atccgcgaagctcgagcaagaagatcaacgcaacttaagtatcccatggtgtg 1230
 |||
 Db 1432 AA 1491
 QY 1231 acgtaccctctgtcttatacagatatacacaagcgtctcaacgagctcatagctcgt 1290
 Db 1492 AA 1551
 QY 1291 gactgataaccctctgattatatacgaagaagacccttaagaatatatcacagacat 1350
 Db 1552 AA 1611
 QY 1351 gagagaagaagttatatacagaagaatcaagaagaatggaagaagaat 1410
 |||
 Db 1612 AA 1671
 QY 1411 gagatgcaagaagaatgacgaagccgcaagaagttcaacgatatcaactaagag 1470
 |||
 Db 1672 AA 1731
 QY 1471 tatgaagaagctgcgaagagatcatgattccaaatcaacataacatgacctgac 1530
 Db 1732 NAAA 1791
 QY 1531 aactcggaagaatgaggaagaacgctaccttcaagaagtggagaacgtgacacacat 1590
 |||
 Db 1792 AA 1851
 QY 1591 aata 1594
 |||
 Db 1852 AAAA 1855
 RESULT 14
 A2548467
 LOCUS
 DEFINITION ENTER30TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
 genomic, DNA sequence.
 ACCESSION A2548467
 VERSION A2548467.1 GI:11172102
 KEYWORDS GSS.
 ORGANISM Entamoeba histolytica.
 Entamoeba histolytica.
 Eukaryota: Entamoebidae; Entamoeba.
 1 (bases 1 to 908)
 REFERENCE Loftus, B., Van Aken, S. and Fraser, C.
 Determination of clone end sequences from Entamoeba histolytica
 HMI:IMSS sheared DNA library
 Unpublished (2000)
 JOURNAL
 COMMENT Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: b.loftus@tigr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Seq primer: M13-Reverse

Query Match 1.4%; Score 68.4; DB 244; Length 596;
 Best Local Similarity 50.3%; Pred. No. 2.8e-06;
 Matches 168; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

```

OY 736 aaaaagaataagaagaccatcgagaacatlaacgagctgtatcgagaatcccaaaagacc 795
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 29 AAGAACCAAGAAAGAACACAGACAGAACAGAACAGAACAGAACAGAACAGAACAG 88
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 796 atagacaaaataagaatgcaaccagaggaagaaagaaagtgtaccagccag 855
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 89 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAACAAA 148
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 856 tacgacctgtccatctataacaacagcttgaagaagccataacctcatcagctactg 915
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 149 AAGAACCAAGAAAGAACACAGACAGAACAGAACAGAACAGAACAGAACAGAACAG 208
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 916 gagaagcgcatagacacccctcaagaagaatgaaataccaagaactgtcgacaagatt 975
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 209 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 268
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 976 aatgaattaaagaatctcctccgccaactctggaacaccccttaacgctgtgac 1035
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 269 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 328
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1036 aagacaagaagaatagaggagcagcagaagaaga 1069
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 329 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 362
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    
```

Search completed: October 27, 2001, 10:44:19
 Job time: 5135 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2001, 13:13:29 ; Search time 34.9 Seconds
(without alignments)
2847.069 Million cell updates/sec

Title: US-09-269-874-3

Perfect score: 8424
Sequence: 1 MKIIFLCFLFIINTQCV.....SNFLGSLFLLMLIYSPI 1639

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_0601:*

- 1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8418	99.9	1639	19	AAW54145
2	8139.5	96.6	1634	6	AAW50777
3	1881	22.3	355	20	AAW09372
4	1881	22.3	355	20	AAW05832
5	1881	22.3	361	20	AAW09373
6	1881	22.3	361	20	AAW05833
7	1869	22.2	376	20	AAW09374
8	1869	22.2	376	20	AAW05834
9	1333	15.8	262	22	AAW37610
10	913	10.8	355	9	AAW80547
11	910.5	10.8	219	9	AAW82523

12	905	10.7	219	9	AAW80544	Polypeptide p190-1
13	794	9.4	244	6	AAW50304	Storage-specific,
14	677	8.0	151	9	AAW80546	Polypeptide p190-2
15	637	7.6	116	18	AAW36103	PfMSP1(p19)A prote
16	637	7.6	116	18	AAW22592	PfMSP1(p19)A prote
17	560	6.6	108	22	AAW37609	Merzoite surface
18	559.5	6.6	1979	21	AAW18171	Plasmodium falcipa
19	556	6.6	96	22	AAW37608	Merzoite surface
20	539	6.4	127	18	AAW22593	PfMSP1(p19)S prote
21	539	6.4	127	18	AAW36102	PfMSP1(p19)S prote
22	455.5	5.4	1254	11	AAW07503	Merzoite apical-en
23	455.5	5.4	1254	18	AAW24575	Merzoite apical-e
24	453.5	5.4	2485	21	AAW18172	Plasmodium falcipa
25	442.5	5.3	3973	21	AAW18253	Plasmodium falcipa
26	431	5.1	980	21	AAW18254	Plasmodium falcipa
27	424	5.0	102	6	AAW50303	Storage-specific,
28	414	4.9	2010	21	AAW18218	Plasmodium falcipa
29	411.5	4.9	1516	21	AAW18195	Plasmodium falcipa
30	408.5	4.8	2539	21	AAW18198	Plasmodium falcipa
31	399	4.7	2013	21	AAW18265	Plasmodium falcipa
32	393	4.7	2954	20	AAW01632	Amino acid sequenc
33	388.5	4.6	2295	21	AAW18180	Plasmodium falcipa
34	387.5	4.6	4134	20	AAW31946	Plasmodium falcipa
35	383.5	4.6	1308	21	AAW18167	Plasmodium falcipa
36	383.5	4.6	2573	21	AAW18234	Plasmodium falcipa
37	379.5	4.5	2206	21	AAW18254	Plasmodium falcipa
38	372	4.4	2500	21	AAW18272	Plasmodium falcipa
39	370	4.4	1411	17	AAW02258	Nucleolar/endosoma
40	369	4.4	1087	20	AAW19935	B. burgdorferi ant
41	369	4.4	1119	20	AAW19934	B. burgdorferi ant
42	365.5	4.3	2380	21	AAW18315	Plasmodium falcipa
43	362.5	4.3	1712	21	AAW18295	Plasmodium falcipa
44	360.5	4.3	1714	21	AAW18275	Plasmodium falcipa
45	357.5	4.2	1817	21	AAW18255	Plasmodium falcipa

ALIGNMENTS

RESULT 1

AAW54145 standard; Protein: 1639 AA.

AAW54145:

23-SEP-1998 (first entry)

P. falciparum synthetic gp190 protein.

gp190: malaria; MSP-1; merozoite surface protein; stability; vaccine;
monoclonal antibody; passive immunisation; parasite.

Plasmodium falciparum.

Synthetic.

W09814583-A2.

09-APR-1998.

02-OCT-1997; 97W0-EP05441.

02-OCT-1996; 96DE-40A0817.

(BUJA/) BUJARD H.

Bujard H, Pan W, Tolle R:

WPI, 1998-240088/21.

N-PSDB: AAV21451, AAV35363.

Recombinant production of complete gp190/MSP-1 Plasmodium surface
protein - useful in anti-malaria vaccines, also stabilising genes by
reducing their AT content

XX
PS. Example 1; Fig 3c; 48pp; German

This sequence represents a modified plasmidium *falciparum* gp190/WSP-1 (merozoite surface) protein. The gene encoding this protein has been stabilised by reducing the AT content of the nucleotide sequence. Such a protein is useful in vaccines against malaria or for producing monoclonal antibodies (for passive immunisation). The complete gp190 protein can now be produced outside the parasite and has, at least over extended regions, the native pattern of folding. Larger amounts of the protein can be produced recombinantly than would be possible using the parasites as source.

SQ Sequence 1639 AA;

Query Match	99.98%	Score 8418;	DB 19;	Length 1639;
Best Local Similarity	99.98%	Pred. No. 0;		
Matches 1638; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0.

Qy	1	MKPEFLCSFPEFIINOCVHESHOEVLTKLELDEAVLITGSLFQKBEKVLINEGSGT	60
Dp	1	mkIifflosstiffinlqcvheshyqelvykkllealeadavltgslfIQekmvlinegtsyt	60
Qy	61	AVTSTPGSKSVASGSGSVASGSGSVASGSGSVASGSGSVASGSGSRRTPNSDNSDS	120
Dp	61	avtsttpsgksvasgsgsgsvasgsgsvasgsgsvasgsgsgsrrtntpsdnsds	120
Qy	121	DAKSYADLKHVRNLTITKELKTPOLFDTNHHMLTCDNHGKTYLIDGEEINELLYK	180
Dp	121	daksyadlkhvrvnyltitkelktpqfdltlnhmltcdnhgfktylidgyeeinellyk	180
Qy	181	LNFEYDLIRAKLNDVCANDYCOIFENLKITRANEDLVAKLIVFGGRKPLDNIKDWGMED	240
Dp	181	lnfydliraklndvcandycqpfnlkitraneldvklivfygrkpldnikdwgmed	240
Qy	241	YIKKKNKTENINELIEBSKTTIDKNKNATKEEBKKLYOAOYDLSITYNQOLEEHNHLS	300
Dp	241	yikkknktieninelieeskrtldknknatkееekkrlyaqydlisitynqlееahnhs	300
Qy	301	VLEKRIOTLKNNENIKELDDINELKNPPRANSNGTPTLTDKKKITEEHKEKELAKT	360
Dp	301	vlekridotlknnenikelddinelknppransngptlldkkkteehkekelakt	360
Qy	361	IKFNIDSLEFDTBLEEYLYREKNNKNIDISAKVEFKESTEPNEHYNGTYPLSYNDINNAL	420
Dp	361	ikfnidsfctdpbleeylyreknknidisakvelkestepnehyngtyplsyndinnal	420
Qy	421	NELNSFGDLINPFDYTKBPSKNITYDNERKKEINEIKETIKERKITESDKSYEDRSKS	480
Dp	421	nelnsfgdlinpfdytkbpskniytdnerkkеinеiketikеrkitesdksyedrskсs	480
Qy	481	LNIDTKEEKLINLEYDSKFNNNIDITNPEEKMMOKRXYVEKXTLHTNTFASVENSNNL	540
Dp	481	lnidtkееklinelеysdkfnniditnpeekmmokrxyvekxtlhtntfasyensnnl	540
Qy	541	EKLTKALYAMDYSHRNIVAVEKEKLKYKNLSKISENETIETIVENIKKDEQOLEEKKRTKD	600
Dp	541	ekltkalyamdyshrnivavekelkyknlskiseNETIETIVENIKKDEQOLEEKKRTKD	600
Qy	601	ENKPEDEKLEVDLVKVOYOKVVLMMNKIDELKQTOLILKNVELKHNTHVPSNRKQENKOE	660
Dp	601	enkpedeklevdsvkvvkvllmmnkidelkqtollknvelkhnthvpsnygenkge	660
Qy	661	PYULYLVITLKEELDKITKVPKPKVESILINEEKKRITRTEGSDNSEPSTBEETIGQATTKRGQO	720
Dp	661	pyulylvitlkeeldkitkvpkpkesilineeKKRITRTEGSDNSEPSTBEETIGQATKpqq	720
Qy	721	AGSALGDSVOAOAOEOKQOAPVPVPVPEPEKAQVPPRPAPVNNKTBNVSKLDVLEKTYE	780
Dp	721	agsalэгdsvgaqgeqdaqpprvpvppekagvpprpapvnnktenvskldyleklye	780
Qy	781	PLMNTSYICHKYLIVSHSTWNEKILKOYKTKITEEBSKLSGCDPLDLFNIONNIPVMSMF	840

Db	761	flnpsjcnkylvshbcmekllkgyk lkeeesklsocdpiddlfnqnmipwysmf	840
Qy	841	DSLNSLSQJLMEIYEREMVONLYKRLKDNK INNLLEAKKYSTSVKTLSSSSMOPLST	900
Db	841	dslnmslsqllmelyekemcnlyklkdncklnlleekavstsvcktlsssmgplst	900
Qy	901	PODKPEVANDDTJSHSNLNSLKLFEENLISLCKMKNYIOELIGKSSNPEYKTLKOSD	960
Db	901	pdkpvevandcteshsnlsmlsklfenllstlqknlygelggkssenfyekllksd	960
Qy	961	TEYNSEFNPFKSKADDINSINDESKKLEEDINKLKTLOISPLVKKYKTLKEPLD	1020
Db	961	ftynesfnfyskaddinsndeskrkleeidinkllktlqstldylknykllkerlfd	1020
Qy	1021	KKKTGKJKMOIKKLTLLKEOLESKLSLNPNPHVLONSVEFNKKKEAIEAENTLEN	1080
Db	1021	kkktvgkjmkgikllktllkeglekslnlnpnhvlgfnfsvfnnkkkeaeiaentlen	1080
Qy	1081	TKLILKHKGLVKVYNGESSPLKTLSEESJOTEDBNYASLENEKVLISKLEGLKLDMLNEK	1140
Db	1081	tklilkhkgjvklyngessplktlseesjtedbnyaslenfkvlisklegklkdnlnlek	1140
Qy	1141	KKLSTLSGCLHHLAELKVTIKKKNYNGNSPNNNDVNNALESYKKPLPECTDVATVS	1200
Db	1141	kllystlsqllhlllaelkevllknnyngnsennclmnalesykkflpegtdvatvs	1200
Qy	1201	ESSGPTLEOSQPKRPASTHGAESNNTTTTSQNDDEVDVYIYPIRGESEEDYDLAGOV	1260
Db	1201	esgspdtleogspkrapasthgaesntttsqnddevdvyilvpirgeeseedydlagov	1260
Qy	1261	TGEAVTPSVDNLSIKLENEYEVLYKPLAGVYRSKLKOLENNVMTFNVNKOILNSRPN	1320
Db	1261	tgeavtpsvdnlsiklenezeyvlykplagvyrsklqlennvmtfnvnkdlnsrfn	1320
Qy	1321	KRENKNVLESDLPYKDILSSNVVWDPKPLNKKRPPKPLSSVNY IKDSIDDPINAN	1380
Db	1321	krenknvlesdldpykdilssnyvwdpklpnkkrrppkplssvnylkdsidctinfn	1380
Qy	1381	DVLGYKYLSEKYSKSDSDSIKKYINDKOGNEKRYLPFLNNIETLYKTVMDITDLEVHLE	1440
Db	1381	dvlgykylsekyksdssdlsikkyindkgenekeylpflnnielyktyndkidtfevhle	1440
Qy	1441	AKVLNYYEKSNVEVKTRELNYLKTTOIDKTRADFKNNNVEGADISTDYNNHNLTKFLS	1500
Db	1441	akvlnyyeksnvevktrelnylkttoktradfknnnfvgadlstdynnhnltkfls	1500
Qy	1501	TGMVPEENAKTVLSNLDDGNLQGLMNTSQHCQYKCCQPNSCGFRLDRECKCKLLWYK	1560
Db	1501	tgmvpennaktvlsnlddgnlqgmlnlsqgcyvckkqcpnsgcftrldreecckllwyk	1560
Qy	1561	OEGBKCVENPPTCENENNNGCCDADAKCTEEDSOSNKKITTCBCTKPDSPYPLDGIJFCSS	1620
Db	1561	gegdckvcpnppctcnennngccdadackcteedsngnkkltcctctkpdspylldgjfcss	1620
Qy	1621	NFLGSLFLILMLILYSFI 1639	
Db	1621	nflgslflilmlilysfi 1639	

RESULT 2
AAP50777
ID AAP50777 standard; Protein; 1654 AA.
XX
XX
AC AAP50777;
XX
DT 30-SEP-1991 (first entry)
XX
XX Sequence of the P195 protein of *Plasmodium falciparum*
DE
XX
XX Malaria vaccine; epitope; antigen; immunogen.
KW
XX
OS *Plasmodium falciparum*.

XX EP154454-A.
XX
XX 11-SEP-1985.
XX
XX 21-FEB-1985; 85EP-0301173.
XX
XX 26-SEP-1984; 84GB-0024340.
XX 22-FEB-1984; 84GB-0004692.
XX 21-FEB-1985; 85GB-0004429.
XX
XX (WELL) WELLCOME FOUNDATION LTD.
XX
XX Holder A. Sandhu J, Odink K, Lockyer M, Riveros-Moreno V;
XX WPI: 1985-224845/37.
XX N-PSDB; AAN50530.
XX
XX Cloned DNA sequence encoding plasmodium falciparum protein -
XX useful for expressing the protein for use in vaccines against
XX malaria
XX
XX Claim 6; Fig 1; 51pp; English.
XX
XX The sequence encoding the P195 protein of Plasmodium falciparum
XX (AAN50530) and a peptide comprising at least one of its epitopes
XX (see AAP50777) are claimed. Also claimed is a vaccine for inducing
XX immunity to malaria comprising the novel peptide or P195 or a
XX peptide comprising at least one epitope when derived from the new
XX DNA sequence, together with a carrier.
XX
XX Sequence 1654 AA:

Query Match 96.6%; Score 8139.5; DB 6; Length 1654;
Best Local Similarity 96.7%; Pred. No. 0;
Matches 1599; Conservative 9; Mismatches 31; Indels 15; Gaps 2;

QY 1 MKIIFFICSPFIIFINOCYTHESYOFELVKKLEALDAVLTGYSLFOKEKAVLNEGTSQT 60
DB 1 MKIIFFICSIIFILINQCVCHESYQELVKKLEALDAVLTGYSLFQKEKAVLNEGTSQT 60
QY 61 AVTTSTGSGKSVASGSGSVASGSGSVASGSGSVASGSGSGNSGRTNPDSNDS 120
DB 61 avtststgsgksvasgsgsvasgsgsvasgsgsvasgsgsgnsrtnpdsnds 120
QY 121 DAKSYADLKHVRNYLTTELKYPQLFDLTNHLTLCDNIGFRKYLIDGYEETNELLYK 180
DB 121 daksyadlkhvrnylttelkypqlfdltfnhltlcdnigfrkylidgyeetnellk 180
QY 181 LNFFEDLRAKLVANDVCQIPFNLRANELDVLKLVFGYRKPLDNTKDNVGMED 240
DB 181 lnffedlraklvandvcqipfnlraneldvlklvfygrkpldntkdnvgmed 240
QY 241 YIKNNKTIENINLEIESKKTIDKNNATPEEEKKKLYQAOQVLSYNNKOLEAHNLIS 300
DB 241 yiknnktieninleieskktidknnatpeekkklyqaqvlsynnkoleahnlis 300
QY 241 YIKNNKTIENINLEIESKKTIDKNNATPEEEKKKLYQAOQVLSYNNKOLEAHNLIS 300
DB 241 yiknnktieninleieskktidknnatpeekkklyqaqvlsynnkoleahnlis 300
QY 301 VLEKRIIDLTKNNENIKELDKINEIKNPPANSNGTPTLLDKNKKIEEHKEIKETAKT 360
DB 301 vlekridltknneinikeldkineiknppansngtptlldknkkieehkeiketakt 360
QY 361 IKFNIDSLFDPLELEYLREKKNKIDISAKVETKESTEPNEYNGVYPLSYNDINNAL 420
DB 361 ikfnidslfdpleleylrekknidisakvetkestepneyngvyplysyndinnal 420
QY 421 NELUSFDLNPDPYKESPKNITDNERKKFIEIKETIKETIKETIKESDOKSYEDSKS 480
DB 421 nelusfdlnpdpkyekspknitdnerkkfieiketikeetikeetikesdoksyedsks 480
QY 481 LNDITTEYKELNLEIVDSKFNNDLITNFERKMGKRYSYKVEKELTHNTFASYSNHNL 540
DB 481 lndittteyekelnleivdskfnndlitnferkmgkrysykvekelthntfasysnhnl 540

QY 541 EKLTKALKMEDYSLRNTVEKEELKYYKNLISKTIENETETLVENIKKDEQLFEKKTITKD 600
DB 541 ekltkalkmedyslrmtnvekeelkyyknlskltienetetlvnikkdeqlfekktitkd 600
QY 601 ENKPDKEILEVSDIVKVOQVLLMNKIDELKKTOLLKKNVLEKHNHIVPVSYSQENKOE 660
DB 601 enkpdekilevsvdivkvovqllmnkidelkktollkknvlekhnhivpvsyseqenke 660
QY 661 PYYLIVLKEIDKLKVPKVESLNEEKKNIKTEGOSDNSEPTGEITGOATTKPGQO 720
DB 661 pyylivlkeidklkvfpkveslneekkniktegosdnseptgeitgoattpgqq 720
QY 721 AGSALLEGDSVQAOQEOQAOAPVPVPVPEAKAQPVPVPVAVNNKTEVNSLDVLEKYE 780
DB 721 agsallegdsvqaqeqqapvpvpvpvpeakaqpvpvpvavnnktevnsldvlekye 780
QY 781 FLNITSYICHKYILVSHSTPMNEKIKQYKITEESKSLSSCDPLDLFNQNNIFVMSMF 840
DB 781 flnitsyichkyilvshstpmnekkikqyikteesksslscdpdlldfnqnnifvmsmf 840
QY 841 DSLN-----NSLSQLEMEIYERKEMVONLYKLDNDKIKNLLEAKKVTSTV 886
DB 841 dslnhvnylwyekelgyvfllimeiyekemvonlykldndkiknlleakkvstsv 900
QY 887 KTLSSSSMQPLSTRPOKPEVSAANDPTSHSTNLNLSKLFENIISLCKNNKIYQELIGQK 946
DB 887 ktlssssmqplstrpoxpevsandptshstnlslsklfeniiislcgnkiyqeligsqk 946
QY 901 KTLSSSSMQPLSTRPOKPEVSAANDPTSHSTNLNLSKLFENIISLCKNNKIYQELIGQK 960
DB 901 ktlssssmqplstrpoxpevsandptshstnlslsklfeniiislcgnkiyqeligsqk 960
QY 947 SSENPYERKILKSDPTFYVESFTFNKSKADINSINDESKKRLKEEDNKKTKTLQSLSD 1006
DB 947 ssenpyeriklksdptfyvesftfnkskaddinsindeskkrllkeednkktktlqslsd 1020
QY 1007 LYNNYKILKLELFDKKTATGYYKMOIKKLLKLEQLSLSLNNPKHVLONFSVFENK 1066
DB 1007 lynnnykilkerlfdkktatgyykmoklllleqlslslnnpkhvlonfsvfennk 1080
QY 1021 LYNNYKILKLELFDKKTATGYYKMOIKKLLKLEQLSLSLNNPKHVLONFSVFENK 1080
DB 1021 lynnnykilkerlfdkktatgyykmoklllleqlslslnnpkhvlonfsvfennk 1100
QY 1067 KEAEIAEENTLNTKILKRYKGLVRYKYNSESSPKITLSEESIOTEDNVASLENFVLS 1126
DB 1067 keaeiaeenlntkilkrykglvrykynseesspkitlseesiotednvaslennfvl 1140
QY 1127 KLEGLKDNLMLEKKKLSYSSGLHHLAELAEVYKNNKNGNSPSENNTPVNNALSES 1186
DB 1141 kleglkdnlmlekkklsysrglhlhlaelaevynknknngnspsenntpvnnalesy 1200
QY 1187 KFLPEGDVATVSESGSDTLEQSPKKPASTHWGAESNTTTTSQNVDEVDVYIIVPIE 1246
DB 1201 kflpegdvatvsegsdtleqspkkpasthwgaesnttttsqnvdevdvyiivlif 1260
QY 1247 GESEEDVDLGOVYVGEAVPVSVIDNLSIKTIENEYEVLYLKLPLAGVRSLSKOLENNVMT 1306
DB 1261 geseedvdlgovvgeavpvsvidnlsiktieneveylylklplagvrsyskolennvmt 1320
QY 1307 FNVNWKDILNRFNKKRENFKNVLESDLIPYKDLTSSNVVDPYKFLNKKERDQFLSSYN 1366
DB 1321 fnvnwkdlnsrfnknrenfnvlesdlipykdltsnvvdpkyflnkerdqflssyn 1380
QY 1367 YIKSDIPTDINFANDVLGYVYLISEKYSKSDLSIKKIYNDKOGENEKYLPLNNIETLYK 1426
DB 1381 yiksdipdinfandvlgvylisekyskdsldskiyndkogenekyplnnietlyk 1440
QY 1427 TVNDKIDLFVTHLEAKVLYNTYKSNVEVKIKELNLYLKTODKLADPKKNNVGIADLS 1486
DB 1441 tvndkidlfvthleakvlyntyksnvevkiikelnylktlqdadfknnvfiadi 1500
QY 1487 TDVNHNNLTLFELSTGWFEFLAKTVLSNLLDGMU-OGMLNISQHOVCVKKQCPNNSGFR 1545
DB 1501 tdvnhnnltrlfelstgwfeelaktvlsnlldgm-ogmlnisohocvkqcpnnsgr 1560
QY 1546 HLDREDECKLLNKKOGSDKCVENPPTCNENNCGCADAKTEDESDSGNKKITCECTK 1605
DB 1561 hldredeckllnykgskskevsnptcnenngcadakteedsngkkitcectk 1620

```

QY 1606 PDSTPLFDGIFCSSSNEFLGIFLLIMLITXSF 1639
DB 1621 pdcyplsmvficssnflglsflllmlllysf 1654

RESULT 3
ID AAY09372 standard; Protein: 355 AA.
AC AAY09372;
DE 31-AUG-1999 (first entry)
DE Merozoite surface protein MSP-1-42.
KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal.
OS Plasmodium falciparum.
XX MO9920774-A2.
XX 29-APR-1999.
XX 20-OCT-1998; 98MO-US222226.
XX 15-MAY-1998; 98US-0085649.
XX 20-OCT-1997; 97US-0062592.
XX (GEN2 ) GENZYME TRANSGENICS CORP.
XX Chen LH, Meade H;
XX WPI: 1999-288313/24.
XX P-PSDB: AAX56008.
XX Modified malarial protein for use in anti-malarial vaccines
XX Example; Fig 1; 35pp; English.
XX The present sequence represents a 42 kDa C-terminal portion of
XX malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
XX important target for the development of a vaccine against
XX Plasmodium falciparum. The nucleic acid sequence encoding MSP-1-42
XX has been modified (see AAX56008) compared to the native sequence (see
XX AAX56009) such that 306 nucleotide positions have been replaced to
XX lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA
XX instability motifs while maintaining the same protein amino acid
XX sequence. These alterations allow MSP-1-42 to be expressed in
XX mammalian cell culture and in transgenic mice. Native MSP-1-12
XX is known to be difficult to express in cell culture systems. The
XX mammalian cell culture systems or in transgenic animals. The
XX invention allows expression of MSP-1 protein in the milk of
XX transgenic animals, and also provides a DNA vaccine comprising a
XX vector containing the altered MSP-1-42 sequence.
XX Sequence 355 AA:
SO

```

```

DB 121 gyyklsekysdsldskkyndckgeneky lplfmllnetcykvndkidlftvhlakv 180
QY 1444 LNTYKSNVEVKIKELNLTCTIDODKADFFKNNNVEGIADLSDYNNMLLTFFLSTGM 1503
DB 181 lntcyksnvevkikelnltctidodkadtffknnnvegiadlscdyhnnmltfflstgm 240
QY 1504 VFENLAKTVLSNLDDGNLQGLNLSOHOCVKKOCPONSGCRHLDERECEKCLNTYKQEG 1563
DB 241 vfenlaktvlsnlddgnlqglnlsgnqcvkkqcgpsgcfrhldereckclntyqeg 300
QY 1564 DKCVENPPTCNENNGCDADAKCTEEDSGSGKRRICETKTPDPSYPLFDGIFCS 1618
DB 301 dkcvenpntcnennngcdadakcteedsngskrricetcktpdspylfdgifs 355

RESULT 4
ID AAY05832 standard; Protein: 355 AA.
AC AAY05832;
DE 02-AUG-1999 (first entry)
DE Merozoite surface protein MSP-1-42.
KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal.
OS Plasmodium falciparum.
XX MO9920766-A2.
XX 29-APR-1999.
XX 20-OCT-1998; 98MO-US222225.
XX 15-MAY-1998; 98US-0085649.
XX 20-OCT-1997; 97US-0062592.
XX (GEN2 ) GENZYME TRANSGENICS CORP.
XX Chen LH, Meade H;
XX WPI: 1999-302742/25.
XX N-PSDB: AAX25586.
XX New modified recombinant nucleic acid sequences useful for producing
XX malarial DNA vaccine
XX Disclosure; Fig 1; 43pp; English.
XX The present sequence represents a 42 kDa C-terminal portion of
XX malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
XX important target for the development of a vaccine against
XX Plasmodium falciparum. The nucleic acid sequence encoding MSP-1-42
XX has been modified (see AAX25586) compared to the native sequence (see
XX AAX25587) such that 306 nucleotide positions have been replaced to
XX lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA
XX instability motifs while maintaining the same protein amino acid
XX sequence. These alterations allow MSP-1-42 to be expressed in
XX mammalian cell culture and in transgenic mice. The invention
XX provides modified recombinant nucleic acid sequences and methods for
XX increasing the mRNA levels and protein expression of proteins that
XX are difficult to express in cell culture systems, mammalian cell
XX culture systems or in transgenic animals. The preferred difficult
XX protein candidates for expression are those derived from lower
XX organisms such as parasites, bacteria and viruses that have DNA
XX coding sequences of high AT content or which have mRNA instability
XX motifs or rare codons relative to the recombinant expression system
XX to be used. The invention allows expression of MSP-1 protein in
XX the milk of transgenic animals, and also provides a DNA vaccine
XX

```

CC comprising a vector containing the altered MSP-1-42 sequence.
 XX Sequence 355 AA;

Query Match 22.3%; Score 1881; DB 20; Length 355;
 Best Local Similarity 100.0%; Pred. No. 1.8e-75;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1264 AVPSVIDNLSKTENEYVLYLKPLAGVYRSKLENNVMTFNVVKDILNSRFNKRRE 1323
 DB 1 avpsvidnlskteniyevevlylkplagvyrsllkqlennvmtfnvkvdlnsrfnkre 60
 QY 1324 NFKNVLESDDLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSSYNTIKSIDTDINPANDVL 1383
 DB 61 nfkvnlesddlipykdltsnnyvvdpykflnkekrdkflssyntiksidtdinfandvl 120
 QY 1384 GYKILSEKYSKSDLSIKKYINDKOGENEKYLPLNNIETLYKTVNDKIDLFVHLEAKV 1443
 DB 121 gyykilssekysdlsikkyindkgenekylplnnietlykcvndkidlfovhleakv 180
 QY 1444 LNTYKESNVEVKIKELNYLKTIDOKLADFRKNNNFVGIADLSTDYHNNLLTFELSTGM 1503
 DB 181 lntyeksnevkiikelnylktldokladfrknnnfvgiadlstdyhnnlltfelstgm 240
 QY 1504 VFENLAKTVLSNLDGMLQGLMNTISOHCYVKKCPONSQCFRHLDERECCCLNTYKOE 1563
 DB 241 vfenlaktvlsnldgmlnsgmqvkkqcpnsgcftrhderceccclntykoe 300
 QY 1564 DKCVENPPTCENNNGGCDADAKCTEEDSGSGKRTCECTRPDSYPLFDGIFCS 1618
 DB 301 dkcvenpntcennnggcdadakcteedsngskrtcectkpsdypldgifcs 355

RESULT 5
 ID AAY09373 standard; Protein: 361 AA.

AC AAY09373;

DT 31-AUG-1999 (first entry)

DE Merozoite surface protein MSP-1-42.

KM MSP-1; merozoite surface protein; malaria; vaccine;
 KM protein engineering; protein expression; codon usage;
 KM transgenic animal.

OS Plasmodium falciparum.

PN WO9920774-A2.

PD 29-APR-1999.

PF 20-OCT-1998; 98WO-US222226.

PR 15-MAY-1998; 98US-0085649.
 PR 20-OCT-1997; 97US-0062592.

PA (GEN2) GENZYME TRANSGENICS CORP.

PI Chen LH, Meade H;

DR WPI; 1999-288313/24.

DR P-PSDB; AAX56009.

Modified malarial protein for use in anti-malarial vaccines
 Example; Fig 2; 35pp; English.

CC This present sequence comprises a 42 kDa C-terminal portion of
 CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
 CC important target for the development of a vaccine against

CC Plasmodium falciparum. The C-terminal end of the sequence is
 CC modified to include a 6xHis tag. A nucleic acid (see AAX56008)
 CC encoding MSP-1-42 has been modified according to a method
 CC of the invention in order to improve expression in mammalian cells
 CC and in transgenic animals by reducing the AT content and removing
 CC mRNA instability motifs. The invention allows expression of
 CC MSP-1-42 in the milk of transgenic animals, and also provides a DNA
 CC vaccine comprising a vector containing the altered MSP-1-42 nucleic
 CC acid.

QY 1264 AVPSVIDNLSKTENEYVLYLKPLAGVYRSKLENNVMTFNVVKDILNSRFNKRRE 1323
 DB 1 avpsvidnlskteniyevevlylkplagvyrsllkqlennvmtfnvkvdlnsrfnkre 60

Query Match 22.3%; Score 1881; DB 20; Length 361;
 Best Local Similarity 100.0%; Pred. No. 1.9e-75;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1264 AVPSVIDNLSKTENEYVLYLKPLAGVYRSKLENNVMTFNVVKDILNSRFNKRRE 1323
 DB 1 avpsvidnlskteniyevevlylkplagvyrsllkqlennvmtfnvkvdlnsrfnkre 60
 QY 1324 NFKNVLESDDLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSSYNTIKSIDTDINPANDVL 1383
 DB 61 nfkvnlesddlipykdltsnnyvvdpykflnkekrdkflssyntiksidtdinfandvl 120
 QY 1384 GYKILSEKYSKSDLSIKKYINDKOGENEKYLPLNNIETLYKTVNDKIDLFVHLEAKV 1443
 DB 121 gyykilssekysdlsikkyindkgenekylplnnietlykcvndkidlfovhleakv 180
 QY 1444 LNTYKESNVEVKIKELNYLKTIDOKLADFRKNNNFVGIADLSTDYHNNLLTFELSTGM 1503
 DB 181 lntyeksnevkiikelnylktldokladfrknnnfvgiadlstdyhnnlltfelstgm 240
 QY 1504 VFENLAKTVLSNLDGMLQGLMNTISOHCYVKKCPONSQCFRHLDERECCCLNTYKOE 1563
 DB 241 vfenlaktvlsnldgmlnsgmqvkkqcpnsgcftrhderceccclntykoe 300
 QY 1564 DKCVENPPTCENNNGGCDADAKCTEEDSGSGKRTCECTRPDSYPLFDGIFCS 1618
 DB 301 dkcvenpntcennnggcdadakcteedsngskrtcectkpsdypldgifcs 355

RESULT 6
 ID AAY05833 standard; Protein: 361 AA.

AC AAY05833;

DT 02-AUG-1999 (first entry)

DE Merozoite surface protein MSP-1-42.

KM MSP-1; merozoite surface protein; malaria; vaccine;
 KM protein engineering; protein expression; codon usage;
 KM transgenic animal.

OS Plasmodium falciparum.

PN WO9920766-A2.

PD 29-APR-1999.

PF 20-OCT-1998; 98WO-US222225.

PR 15-MAY-1998; 98US-0085649.
 PR 20-OCT-1997; 97US-0062592.

PA (GEN2) GENZYME TRANSGENICS CORP.

PI Chen LH, Meade H;

DR WPI; 1999-302742/25.

DR N-PSDB; AAX25587.

Query Match	Best Local Similarity	Score	DB	Length
Matches 353; Conservative	99.4%	22.2%	1869	376
0; Mismatches	0; Indels	0; Gaps		

QY	1264	AVTSVSDNLTSLKENEYEVLYLTKPLAGVYRSJLKKOLENNWFENWVKDILNSFKRKE	1323
Db	16	avcpvsidnltslkieniyeewlylkplagvyrslkqjlenmvntfnvkvdlinsfnkre	75
QY	1324	NFKVWLESDDLPRKDDLTSSWVVKDPRKFLNKKRDPFLSSVNYIKDSIDTDINFDVNL	1383
Db	76	nfkvlaseddlprkddltssnyvvvkdkprrkflnkkrdkflssnylksdldcdinfandvl	135
QY	1384	GYYKILSEKYSDDLSIKKYINDKQGENEKYLPFLNNIETLTKYTNXKIDLFVHLKAY	1443
Db	136	gyyllsekyksdldslkkyindkqgenekyrlpflnnietlyktyvnxkidlflvhlleakv	195
QY	1444	LNITYEESNVEVKIKELNYTKTQDKLAPFKNNNFVGADLSDTYNNHNLTKFLSGM	1503
Db	196	lgytyeksnvevkkikelnylktlqdkladrkfknnfnvgiadlstqymhnlitkflstqgm	255
QY	1504	VFEKLAQTVLSNLLDGLMGLMISQHOVCVKKOCPONSGCFPHLDERECCCLNMYKQEG	1563
Db	256	vfehlaktvlsnllldglmgmlisqhqcvkkyqcpnsgcfphlderecccllnmykqeg	315
QY	1564	DKCEVNPNCNENNGCCDADAKCTEEDSGSNGKKITCCTKPDSPYPLFDGIFCS	1618
Db	316	dkcevpnpncnenngccdadactceedsngskkitcctkpdspylfdgifcs	370
RESULT 9			
AAAB37610	AAAB37610 standard; Protein: 262 AA.		
AC	AAAB37610:		
DT	27-FEB-2001 (first entry)		
DE	Merozoite surface protein-133.		
KM	Merozoite surface protein; protazoacide; vaccine; malaria.		
OS	Plasmodium falciparum.		
PN	W0200063245-A2.		
PD	26-OCT-2000.		
PF	20-APR-2000; 2000WO-GB01558.		
PR	20-APR-1999; 99GB-0009072.		
PR	13-MAY-1999; 99US-0311817.		
PR	25-MAY-1999; 99CA-2271451.		
PA	(MED-) MEDICAL RES COUNCIL.		
PI	Holder A, Birdsell B, Feeney J, Morgan W, Syed S, Uthaiyibull C;		
DR	WPI: 2001-015762/02.		
DR	N-PSDB; AAC68978.		
PT	Novel variants of the C-terminal fragment of Plasmodium merozoite		
PT	surface protein-1, useful as vaccines for treating or preventing		
PT	malaria		
XX	Example 5; Fig 15; 126pp; English.		
XX	The present invention relates to non-natural variants of a C-terminal		
XX	fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The		
XX	non-natural variants have reduced affinity for at least 1 antibody		
XX	capable of blocking a second antibody that inhibits the proteolytic		
XX	cleavage of Plasmodium MSP-1-4.2, and has the same affinity for at least		
XX	one third antibody that inhibits the proteolytic cleavage of Plasmodium		
XX	MSP-1-4.2, compared to natural MSP-1-1.9. The non-natural variants of the		
XX	present invention are useful for immunising a mammal against malaria, and		
XX	can be used to treat malaria. The present sequence is MSP-133 protein.		
XX	Sequence 262 AA:		

Query Match 15.8% Score 1333; DB 22; Length 262;
 Best Local Similarity 99.6%; Pred. No. 1.3e-51;
 Matches 261; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1264 AVPPSVIDNLSIKTENEXEVLTKPLAGVRSRLKROLENNVTFNVNKKDILNSRFNRE 1323
 1 avtpsvidnlsiktenexevltpkplagvrsrlkqleennvtfnvknkdlnsrfnre 60
 Db 1264 AVPPSVIDNLSIKTENEXEVLTKPLAGVRSRLKROLENNVTFNVNKKDILNSRFNRE 1323
 1 avtpsvidnlsiktenexevltpkplagvrsrlkqleennvtfnvknkdlnsrfnre 60
 QY 1324 NFKVNLSDLPYCDLTISSNNVVKDPYKFLNKEKRDKFLSSYNTIKOSIDPDINADV 1383
 61 nfkvnlsdldpycdltissnnyvkdpykflnkekrdkflssyntikdsidtdinadv 120
 Db 1384 GYVILSEKYSKSDSDSTIKKYINDKOGENEKYLPLNNETLYKTWVNDKIDLFVHLEAKV 1443
 121 gyvilseskyskdsdstikkyindkgenekyplnnetylktwvndkidlfvhlakv 180
 QY 1444 LNTYERKSNVEKIKELNYLTQIDKLADFKKNNNFVGIAIDSTDYNNHNLTRFLSTGM 1503
 181 lntyersnvevkiikelnyltqidkladfkknfnvgiadlstdynhnltrflstgm 240
 Db 1504 VFENLAKTVLSNLDGNLQGM 1525
 241 vfenlaktvlsnldgnlqgm 262

RESULT 10
 AAP80547 ID AAP80547 standard; protein: 355 AA.
 AC AAP80547;
 XX
 DT 11-NOV-1990 (first entry)
 XX
 DE Polypeptide p190-3.
 XX
 KM Polypeptide p190-3; P. falciparum; merozoite; vaccine.
 XX
 OS Plasmodium falciparum.
 XX
 FH Key Location/Qualifiers
 FT Region 1..7
 FT /label=affinity peptide
 FT Region 144..355
 XX
 PN EP283829-A.
 XX
 PD 28-SEP-1988.
 XX
 PE 08-MAR-1988; 88EP-0103564.
 XX
 PR 19-MAR-1987; 87GB-0006599.
 XX
 PA (HOFF) F. HOFFMANN-LA ROCHE & CO.
 XX
 PI Ulrich C, Gentz H, Takacs B;
 XX
 DR WPI; 1988-272339/39.
 DR N-PSDB; AAN81151.
 XX
 PT New immunogenic polypeptides - derived from Plasmodium falciparum 190 kD
 PT surface antigen precursor, useful in vaccines, and encoding DNA sequences
 XX
 PS Claim 7, Page 28; 45pp; German.
 XX
 CC The polypeptide contains at least one epitope of the 190 kD precursor of
 CC the major merozoite surface antigen of P. falciparum. The polypeptide is
 CC immunogenic and can be used in vaccines to protect against malaria.
 CC Presence of the affinity peptide allows the polypeptide to be purified
 CC by affinity chromatography resins. Antibodies raised against the
 CC polypeptide are useful for diagnosis and for purification by affinity
 CC chromatography.

XX
 SQ Sequence 355 AA;

Query Match 10.8% Score 913; DB 9; Length 355;
 Best Local Similarity 80.6%; Pred. No. 5e-33;
 Matches 187; Conservative 9; Mismatches 14; Indels 22; Gaps 4;

QY 105 SGNSRRNPSSNDSOAKSYADLKHRRNRYLLTIKELKYQLEFDLNNHML-----TLCD 159
 106 sgnsrrnpssnssdsaksyadlkhrrnrylltikelkyqlefdlnnhml-----tlcd 148
 Db 106 tgn-----psvnnnd-----vnalesykhf-lpegdvalvsvessgqlcd 148
 QY 160 NHHGFKYLIDYEEINELLYKINFPDLIRAKLNDVANCQIPFNKIRANLADYLK 219
 149 nhhgfkylidgyeeinellykintfydliiraklnvncandycqipfnkiraneldylk 208
 Db 220 LVFGYRPLDNKKNVGMEDYIKKNNKTENINELIBESKRTIDKKNATKEEKKKLY 279
 209 lvfgyrpldnkknvghmedyiknnkkteninellieskrtldkknatkkeekkkly 268
 QY 280 QAOYDLSTYNNKOLEAHNLISVLEKRRIDTLKKNENIKELLDKINERPNPPA 331
 269 qaydliynkqleahnlisvlekrridtlkknenikelldkineirnpnpa 320
 Db

RESULT 11
 AAP82523 ID AAP82523 standard; protein: 219 AA.
 AC AAP82523;
 XX
 DT 12-DEC-1990 (first entry)
 XX
 DE 31-1 Repeated Delete.
 XX
 KM Malaria; stage-specific late schizont-merozoite antigen 31-1;
 KM vaccine; protozoan parasite; reptile.
 XX
 OS Plasmodium falciparum.
 XX
 PN EP254862-A.
 XX
 PD 03-FEB-1988.
 XX
 PE 20-JUN-1987; 87EP-0108867.
 XX
 PR 26-JUN-1986; 86US-0879076.
 XX
 PA (BEHW) BEHRINGWERKE AG.
 XX
 PI Shaw A, Humbert Y;
 XX
 DR WPI; 1988-030152/05.
 DR N-PSDB; AAN82176.
 XX
 PT Peptide(s) for prodn. of antiprotozoal vaccines - comprising
 PT surface antigen sequence with deleted reptile
 XX
 PS Disclosure; ; p: English.
 XX
 CC The sequence encodes a polypeptide which contains the unique
 CC sequence of the surface antigen of P. falciparum, but lacks the
 CC immunodominant repeat sequences which may allow the parasite to
 CC evade and decoy the immune system. The DNA for the stage-
 CC specific late schizont-merozoite antigen 31-1 in p31-1 (W08503725)
 CC was modified to delete the entire sequence coding for the reptile.
 CC The expression plasmid, 31-1 Repeated Delete, was introduced into
 CC E. coli K12 to produce the modified peptide. The transformant has
 CC been deposited in the Deutsche Sammlung von Mikroorganismen as
 CC E. coli K12 (958336c1) (p31-1 Repeat Delete). The peptide was used
 CC in compns. to treat and prevent protozoan parasitic infections.
 XX
 SQ Sequence 219 AA;

DB 189 daksyadlkhrvqnylflkelkypelldlenhnlldcnlnhgfkylldgyeei 242

RESULT 14

AAW36103 ID AAP80546 standard; protein: 151 AA.

XX AAP80546;

DT 11-NOV-1990 (first entry)

DE Polypeptide p190-2b.

XX Polypeptide p190-2b; P.falciparum; merozoite; vaccine.

OS Plasmodium falciparum.

XX Key Location/Qualifiers

FT Region 1..5

FT Region /label=affinity peptide

PN EP283829-A.

PD 28-SEP-1988.

PF 08-MAR-1988; 88EP-0103564.

PR 19-MAR-1987; 87GB-0006599.

PA (HOFF) F. HOFFMANN-LA ROCHE & CO.

PI Ulrich C, Gentz H, Takacs B;

DR WPI: 1988-272339/39.

DR N-PSDB; AAN81150.

PT New immunogenic polypeptides - derived from plasmodium falciparum 190 kd

PS Claim 6; Page 27; 45pp; German.

CC the polypeptide contains at least one epitope of the 190 kd precursor of

CC the major merozoite surface antigen of P.falciparum. The polypeptide is

CC immunogenic and can be used in vaccines to protect against malaria.

CC Presence of the affinity peptide allows the polypeptide to be purified

CC by affinity chromatography resins. Antibodies raised against the

CC polypeptide are useful for diagnosis and for purification. by affinity

CC chromatography.

CC Sequence 151 AA;

QY 1069 AEIETENTLTKLILKHYGLVYNGSSPLKLTSEESIQEDNYASLENFKVLSKL 1128

DB 6 aeietentlentklllkhyglvkyngessplkltseesiqednyaslenfkvlskl 65

QY 1129 EGKJLKDNLNLEKKKLSYLSGLHLLIAELKEVIRKNKNTGNSPENNNTDNNALSYKRF 1188

DB 66 egkjdknlnlekkklsysrglhhlliaelkevirknkntgnspsennntdnnalasykrf 125

QY 1189 LPEGTDAVATVVSSEGSSTLEQS 1210

DB 126 lpegtvatvvsessgsstlqps 147

RESULT 15

AAW36103 ID AAW36103 standard; protein: 116 AA.

XX AAW36103;

AC 25-MAR-1998 (first entry)

DE P1MSPI(p19)A protein sequence.

XX Plasmodium vivax; merozoite surface protein; MSP1; p19;

XX Plasmodium falciparum; malaria; vaccine; immunity; epitope.

OS Plasmodium falciparum.

XX Synthetic.

XX Key Location/Qualifiers

FT Region 1..95

FT Region /note="amino acids derived from P. falciparum MSP1 p19

FT Region /note="glycosylphosphatidylinositol anchoring sequence"

PN W09730158-A2.

PD 21-AUG-1997.

PF 14-FEB-1997; 97WO-FR00290.

PR 14-FEB-1996; 96FR-0001822.

PA (INSP) INST PASTEUR.

PI (UNY) UNIV NEW YORK STATE.

PI Barnwell JW, Longacre-Andre S, Mendis K, Nato F;

DR WPI: 1997-425033/39.

DR N-PSDB; AAT94550.

PT Recombinant protein containing the merozoite surface protein-1 p19

PT fragment - useful in anti-malarial vaccines, diagnosis and protein

PT purification

PS Disclosure; Fig 1B; 85pp; French.

XX This is the amino acid sequence of a recombinant protein comprising

XX amino acids 1613-1705 of the Plasmodium falciparum merozoite surface

XX protein 1 (MSP1) 19 kd C-terminal fragment (p19). Linked to a

XX glycosylphosphatidylinositol membrane anchoring sequence. p19 is the

XX C-terminal fragment of the 42 kd MSP1 from Plasmodium species.

XX The recombinant protein can be used for the production of anti-malarial

XX vaccines, where the p19 fragment provides a high level of protective

XX immunity since it includes epitopes not presented in the p42 fragment.

XX Sequence 116 AA;

QY 1526 NISOHCYKQCPONSGCFRLHDERECKCLNKKQESDKCVENPNPTGNNNGCCDADA 1585

DB 3 nlsqgcvkkqcpensgcfrrlhderECKclnykqegdkcvENpnpTCNENNGCCDADA 62

QY 1586 KCTERDSSNKKRTTCECTKPDSPPLPFGITCCSSNPLGISTFLIMLILXSF 1639

DB 63 kctedsngskktlcectkpdspyltfgitccssnplgistsflimlilysfi 116

Search completed: October 27, 2001, 15:51:14

Job time: 9465 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2001, 15:51:48 ; Search time 379.17 Seconds
(without alignments)
8180.589 Million cell updates/sec

Title: US-09-269-874-2

Perfect score: 4940

Sequence: 1 ccgacgcgcatgaataatcat.....ttcatctaatagatgatg 4940

Scoring table: IDENTITY_NNC
Gapop 10.0 , Gapext 1.0

Searched: 730101 segs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_0601: *
1: /SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT: *
2: /SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT: *
3: /SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT: *
4: /SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT: *
5: /SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT: *
6: /SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT: *
7: /SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT: *
8: /SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT: *
9: /SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT: *
10: /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT: *
11: /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT: *
12: /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT: *
13: /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT: *
14: /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT: *
15: /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT: *
16: /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT: *
17: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT: *
18: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT: *
19: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT: *
20: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT: *
21: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT: *
22: /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4938.4	100.0	4940	19	AAV21451
2	2837.6	57.4	4940	19	AAV5363
3	2706.2	54.8	5760	6	AAV50530
4	1262	25.5	5181	16	AAO80911
5	1000	20.2	1950	18	AAV97956
6	989	20.0	1897	18	AAV97958
7	969.6	19.6	1896	18	AAV97957
8	744.2	15.1	1065	20	AAV56008
9	744.2	15.1	1065	20	AAV5586
10	742	15.0	1140	20	AAV56021
11	742	15.0	1140	20	AAV5593

12	719.8	14.6	1077	22	AAV68976	Merozoite surface
13	625.2	12.7	1088	20	AAV56009	Merozoite surface
14	625.2	12.7	1088	20	AAV5587	Merozoite surface
15	525.2	10.6	786	22	AAV68978	Merozoite surface
16	308.8	6.3	1068	9	AAV81148	DNA sequence encod
17	308.8	6.3	1068	9	AAV81151	DNA sequence encod
18	246.6	5.0	354	18	AAV80403	PFMSPI(P19)A codin
19	246.6	5.0	354	18	AAV94550	PFMSPI(P19)A codin
20	243.4	4.9	737	6	AAV50355	Storage-specific,
21	240.2	4.9	668	9	AAV82176	31-1 Repeated dele
22	201.8	4.1	456	9	AAV81150	DNA sequence encod
23	201.8	4.1	462	9	AAV81149	Merozoite surface
24	198.4	4.0	333	22	AAV68977	PFMSPI(P19)S codin
25	194.8	3.9	387	18	AAV80404	PFMSPI(P19)S codin
26	194.8	3.9	387	18	AAV94549	PFMSPI(P19)S codin
27	126.2	2.6	936	22	AAV58252	Oligonucleotide D1
28	126.2	2.6	936	22	AAV58254	Oligonucleotide D1
29	126.2	2.6	936	22	AAV58257	Oligonucleotide D1
30	126.2	2.6	936	22	AAV58259	Oligonucleotide D2
31	126.2	2.6	936	22	AAV58262	Oligonucleotide D2
32	126.2	2.6	938	22	AAV58255	Oligonucleotide D1
33	122.6	2.5	936	22	AAV58252	Oligonucleotide D1
34	122.6	2.5	936	22	AAV58254	Oligonucleotide D1
35	122.6	2.5	936	22	AAV58257	Oligonucleotide D1
36	122.6	2.5	936	22	AAV58259	Oligonucleotide D2
37	122.6	2.5	936	22	AAV58262	Oligonucleotide D2
38	122.6	2.5	938	22	AAV58255	Oligonucleotide D1
39	119.2	2.4	306	6	AAV50354	Storage-specific,
40	105.6	2.1	3399	17	AAV05868	Chicken leucocytos
41	99.4	2.0	165	18	AAV93729	DNA encoding signa
42	81.6	1.7	1686	16	AAV87587	DNA encoding leuco
43	70.6	1.4	3579	21	AAV70099	Plasmodium falci
44	65.2	1.3	8045	21	AAV55171	Neurospora crassa
45	60.6	1.2	5163	19	AAV20700	Cryptosporidium pa

ALIGNMENTS

RESULT 1	
AAV21451	standard; DNA: 4940 BP.
ID	AAV21451
XX	AAV21451
AC	AAV21451
DT	23-SEP-1998 (first entry)
DE	P. falciptarum modified gp190 DNA.
XX	
KW	gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
KW	monoclonal antibody; passive immunisation; parasite; ss.
XX	
OS	Plasmodium falciptarum.
OS	Synthetic.
XX	
FT	Key
FT	CDS
FT	location/qualifiers
FT	/*tag= a
FT	/product= gp190
XX	
PD	W09814583-A2.
XX	
PD	09-APR-1998.
XX	
PF	02-OCT-1997; 97W0-EP05441.
XX	
PR	02-OCT-1996; 96DE-4040817.
XX	
PA	(BUA/) BUARD H.
XX	
PI	Bujard H, Pan W, Tolle R;
XX	
DR	WPI; 1998-240088/21.

Db 1801 acaaaagcgaataaacaacagatgagagatcctgagctccgatatgttgaagtc 1860
Qy 1861 caagtgccgaaggggtgctctctatgatacaagaattgatgaaccagaagactcaatctt 1920
Db 1861 caagtgccgaaggggtgctctctatgatacaagaattgatgaaccagaagactcaatctt 1920
Qy 1921 ctggaagacggtgaggtttaaacaataatacattgctgcgaatagttaagcagagat 1980
Db 1921 ctggaagacggtgaggtttaaacaataatacattgctgcgaatagttaagcagagat 1980
Qy 1981 aagcaggaacatactactcctacgttactcaagaagaagatagacaaactgaagtgttc 2040
Db 1981 aagcaggaacatactactcctacgttactcaagaagaagatagacaaactgaagtgttc 2040
Qy 2041 atgcccgaagtcgagagcttgatcaacgaagaagaagaacatataaactggaagcag 2100
Db 2041 atgcccgaagtcgagagcttgatcaacgaagaagaagaacatataaactggaagcag 2100
Qy 2101 tcagataaactccgagccttccacagaagaagataaaccggaacggtctacccaagccc 2160
Db 2101 tcagataaactccgagccttccacagaagaagataaaccggaacggtctacccaagccc 2160
Qy 2161 ggacaacacgcccgttccagctctcgaagcgatagcgtgcaagctcaagcacaagacag 2220
Db 2161 ggacaacacgcccgttccagctctcgaagcgatagcgtgcaagctcaagcacaagacag 2220
Qy 2221 aagcagacacgctccagctccagctgcccgttccagaggttaagctcaagtgctaca 2280
Db 2221 aagcagacacgctccagctccagctgcccgttccagaggttaagctcaagtgctaca 2280
Qy 2281 ccacacgctcctgtgaataaacaagccgagaatgtccagcaactggaactcttgaag 2340
Db 2281 ccacacgctcctgtgaataaacaagccgagaatgtccagcaactggaactcttgaag 2340
Qy 2341 ctctatgagttccctgaatacatcctacactgtccacaataatatacctgtctctcagc 2400
Db 2341 ctctatgagttccctgaatacatcctacactgtccacaataatatacctgtctctcagc 2400
Qy 2401 actatgaacggaagaattcttaaacagtaacagataacccaaggaagagataaactgtg 2460
Db 2401 actatgaacggaagaattcttaaacagtaacagataacccaaggaagagataaactgtg 2460
Qy 2461 tccctctgtgataccacttgaacctgtgttcaatataccagaaacaactccggttattgt 2520
Db 2461 tccctctgtgataccacttgaacctgtgttcaatataccagaaacaactccggttattgt 2520
Qy 2521 tctatgtctgataagcctcaaatctctctctcaactgttcaatgagagataatagaag 2580
Db 2521 tctatgtctgataagcctcaaatctctctctcaactgttcaatgagagataatagaag 2580
Qy 2581 gagaatggtctgaacctgttataaactcaaaagcaacgaagaattaaagaactcttgag 2640
Db 2581 gagaatggtctgaacctgttataaactcaaaagcaacgaagaattaaagaactcttgag 2640
Qy 2641 gaaactgaagaagttctccaactctgtttaaaactctctctccagcttccagcaactgtg 2700
Db 2641 gaaactgaagaagttctccaactctgtttaaaactctctctccagcttccagcaactgtg 2700
Qy 2701 tctctcacacctcaagacagcccgaaagtgaagcgcttaacgagacactctcaactcgacc 2760
Db 2701 tctctcacacctcaagacagcccgaaagtgaagcgcttaacgagacactctcaactcgacc 2760
Qy 2761 aaccttaataactcaactgaactgtttgagaacatctcgtctctcgggaagaataagac 2820
Db 2761 aaccttaataactcaactgaactgtttgagaacatctcgtctctcgggaagaataagac 2820
Qy 2821 atctcaacgaacttatgtgacagaacatcgctccgagaactctcaaggaagatactgaaa 2880
Db 2821 atctcaacgaacttatgtgacagaacatcgctccgagaactctcaaggaagatactgaaa 2880
Qy 2881 gacagcgacacatctataacgagagcttcaactctgtgaaatctaaagccgagt 2940
Db 2881 gacagcgacacatctataacgagagcttcaactctgtgaaatctaaagccgagt 2940

Qy 2941 atcaactctcttaagatgaatcttaaacgtlaagaagcgggaagagacataaagctg 3000
Db 2941 atcaactctcttaagatgaatcttaaacgtlaagaagcgggaagagacataaagctg 3000
Qy 3001 aagaagacactgcaactgaactgactgacactgttacaacaagctcaactgaactgagaga 3060
Db 3001 aagaagacactgcaactgaactgactgacactgttacaacaagctcaactgaactgagaga 3060
Qy 3061 ctcttcgaagaagaagacagctcggaagtataagatgacatcaagaagtgtgactg 3120
Db 3061 ctcttcgaagaagaagacagctcggaagtataagatgacatcaagaagtgtgactg 3120
Qy 3121 ctcaagagcagcttgaaagaacactcaactcactgacacatctcgaacagatctgacg 3180
Db 3121 ctcaagagcagcttgaaagaacactcaactcactgacacatctcgaacagatctgacg 3180
Qy 3181 aacttctcagtgcttctcaacaagaagaagagcggagatccgacgaagacagacact 3240
Db 3181 aacttctcagtgcttctcaacaagaagaagagcggagatccgacgaagacagacact 3240
Qy 3241 ctggaagaacccaagaattctctcaacaacactacaagaagcctcgtaagtattataatg 3300
Db 3241 ctggaagaacccaagaattctctcaacaacactacaagaagcctcgtaagtattataatg 3300
Qy 3301 gacttctctcttgaagaactctctcgaagagagcagcagcagcagatcaagc 3360
Db 3301 gacttctctcttgaagaactctctcgaagagagcagcagcagcagatcaagc 3360
Qy 3361 agcctcgaagacttcaagaactctctgctlaagctcggaagcagcgtlaagacacactgac 3420
Db 3361 agcctcgaagacttcaagaactctctgctlaagctcggaagcagcgtlaagacacactgac 3420
Qy 3421 ctggaagaagaagaagctcagctacactctctagcgaactgacatcgacgagc 3480
Db 3421 ctggaagaagaagaagctcagctacactctctagcgaactgacatcgacgagc 3480
Qy 3481 aaggaagtcataaagaacaagaactacacccgcaatagcccaagcggaataataacag 3540
Db 3481 aaggaagtcataaagaacaagaactacacccgcaatagcccaagcggaataataacag 3540
Qy 3541 gtgaataacgcaactggaattctatacaagaagtctctgacttgaagaacagatgctgcact 3600
Db 3541 gtgaataacgcaactggaattctatacaagaagtctctgacttgaagaacagatgctgcact 3600
Qy 3601 gtggtgtctgaatctggtctcgacacacttgaagcagcttcaactaaagaagcgtgacact 3660
Db 3601 gtggtgtctgaatctggtctcgacacacttgaagcagcttcaactaaagaagcgtgacact 3660
Qy 3661 actcagtcgagccgagctccaatacaattacccaactcgaagaagctgacagatgagct 3720
Db 3661 actcagtcgagccgagctccaatacaattacccaactcgaagaagctgacagatgagct 3720
Qy 3721 gatgacgtcaltcaltgtgctcattctcgcgagagccgagggagactagatgacctgcg 3780
Db 3721 gatgacgtcaltcaltgtgctcattctcgcgagagccgagggagactagatgacctgcg 3780
Qy 3781 caggtgttcaacccgttgagcgtgcaactccttcggtgattgataaactctgtccaactc 3840
Db 3781 caggtgttcaacccgttgagcgtgcaactccttcggtgattgataaactctgtccaactc 3840
Qy 3841 gagaagcaatacgaagtgtctctatctgaacactcgggaagcggttataggtcttcag 3900
Db 3841 gagaagcaatacgaagtgtctctatctgaacactcgggaagcggttataggtcttcag 3900
Qy 3901 aaacagctggaagaataacgtatgacacttcaatgtccaagtgaagacatctgacagc 3960
Db 3901 aaacagctggaagaataacgtatgacacttcaatgtccaagtgaagacatctgacagc 3960
Qy 3961 cgtttaaagaagaagaatttcaagaagcttlyggaagcgagctgattccctataa 4020
Db 3961 cgtttaaagaagaagaatttcaagaagcttlyggaagcgagctgattccctataa 4020

```

QY 4021 gacctgacctctctaactagctgttcaaggaccatacaagtctcctaataagaagaag 4080
    |||||||
DB 4021 gacctgacctctctaactagctgttcaaggaccatacaagtctcctaataagaagaag 4080
QY 4081 agggataaattctctagttacaactatatacaagatccatcgacaacgatacaat 4140
    |||||||
DB 4081 agggataaattctctagttacaactatatacaagatccatcgacaacgatacaat 4140
QY 4141 ttccgtaattgtctgtgggtattacaagaatcctgaagcggaaataataaagtctgaact 4200
    |||||||
DB 4141 ttccgtaattgtctgtgggtattacaagaatcctgaagcggaaataataaagtctgaact 4200
QY 4201 gactctataaataaataatatacaacgataaagcgaagcgaatgtaaaatattcgccttc 4260
    |||||||
DB 4201 gactctataaataaataatatacaacgataaagcgaagcgaatgtaaaatattcgccttc 4260
QY 4261 ctgaataacatcgaaacccctgtacagaacagtgaacgacaatatcgacctcttgtaatt 4320
    |||||||
DB 4261 ctgaataacatcgaaacccctgtacagaacagtgaacgacaatatcgacctcttgtaatt 4320
QY 4321 cacctggagggccaaagctctcaactactactacgagaagagagaatgtggaatttaaaatc 4380
    |||||||
DB 4321 cacctggagggccaaagctctcaactactactacgagaagagagaatgtggaatttaaaatc 4380
QY 4381 aaggagctgaactactactcaaaacaaatccaaagacaagctgacgaattccaagaataaac 4440
    |||||||
DB 4381 aaggagctgaactactactcaaaacaaatccaaagacaagctgacgaattccaagaataaac 4440
QY 4441 aatttcgttcggaatttgacgaactgtctacacgattataacacacaatctcctgaccaa 4500
    |||||||
DB 4441 aatttcgttcggaatttgacgaactgtctacacgattataacacacaatctcctgaccaa 4500
QY 4501 ttctgtctcactgacgtgtgtttcgaaaacccctgcgcaaaacagtgctgagcaattgtctc 4560
    |||||||
DB 4501 ttctgtctcactgacgtgtgtttcgaaaacccctgcgcaaaacagtgctgagcaattgtctc 4560
QY 4561 gacggcaaacctgacgaggtacgtgacacatctccacgacaaatgtgtgaagaacaagtgc 4620
    |||||||
DB 4561 gacggcaaacctgacgaggtacgtgacacatctccacgacaaatgtgtgaagaacaagtgc 4620
QY 4621 ccccaagaataagcgtgtttcgaagcatctgagacgacggaagatgcaagtgtctctg 4680
    |||||||
DB 4621 ccccaagaataagcgtgtttcgaagcatctgagacgacggaagatgcaagtgtctctg 4680
QY 4681 aactacaacaagaagagatgaatgctgtgagaacccaacccctacctgcaatgaataaac 4740
    |||||||
DB 4681 aactacaacaagaagagatgaatgctgtgagaacccaacccctacctgcaatgaataaac 4740
QY 4741 aatggcggtgtgtgacgcgcatgctaataatgcacccgaggaagacgagctctaagcgaag 4800
    |||||||
DB 4741 aatggcggtgtgtgacgcgcatgctaataatgcacccgaggaagacgagctctaagcgaag 4800
QY 4801 aaatcacatcgaggtgacgaagccgacacccctctcttaacatttgaagggattttttgc 4860
    |||||||
DB 4801 aaatcacatcgaggtgacgaagccgacacccctctcttaacatttgaagggattttttgc 4860
QY 4861 tcacgcttaattctcctgacatccctctcctgctgacatccatgatacctgtgacaagc 4920
    |||||||
DB 4861 tcacgcttaattctcctgacatccctctcctgctgacatccatgatacctgtgacaagc 4920
QY 4921 ttcatctaataagatcatgagc 4940
    |||||||
DB 4921 ttcatctaataagatcatgagc 4940

```

RESULT 2
AAV35363
ID AAV35363 standard; DNA; 4940 BP.

XX AAV35363;
XX
DT 23-SEP-1998 (first entry)
XX

```

DE P. falciparum gp190 DNA.
XX
KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
KW monoclonal antibody; passive immunisation; parasite; ss.
XX
OS Plasmodium falciparum.
XX
FH Key location/Qualifiers
FT CDS 10..4929
FT /tag=a
FT /product= gp190
XX
XX MO9814583-A2.
XX
XX PD 09-APR-1998.
XX
XX PF 02-OCT-1997; 97WO-EP05441.
XX
XX PR 02-OCT-1996; 96DE-4040817.
XX
XX PA (BUA/) BUJARD H.
XX
XX PI Bujard H, Pan W, Tolle R.
XX
XX DR WPI; 1998-240088/21.
XX
XX DR P-PSDB; AAW54145.
XX
XX PT Recombinant production of complete gp190/MSP-1 Plasmodium surface
PT protein - useful in anti-malaria vaccines, also stabilising genes by
PT reducing their AT content
XX
XX PS Example 1; Fig 3c; 48pp; German.
XX
XX CC This sequence encodes the Plasmodium falciparum gp190/MSP-1 (merozoite
CC surface protein). This gene is used in a method for stabilising the
CC gene sequences by reducing the AT content. Such products are useful in
CC vaccines against malaria or for producing monoclonal antibodies (for
CC passive immunisation). The complete gp190 protein can now be produced
CC outside the parasite and has, at least over extended regions, the native
CC pattern of folding. Larger amounts of the protein can be produced
CC recombinantly than would be possible using the parasites as source.
XX
XX SQ Sequence 4940 BP; 2196 A; 597 C; 687 G; 1460 T; 0 other;

Query Match 57.4%; Score 2837.6; DB 19; Length 4940;
Best Local Similarity 73.4%; Pred. NO. 0; Mismatches 1314; Indels 0; Gaps 0;
Matches 3626; Conservative 0;

QY 1 cgaacgcatgaataatcttctcctcgttcattctgttttatacataact 60
    |||||||
DB 1 cgaacgcatgaataatcttctcctcgttcattctgttttatacataact 60
QY 61 cagtcgtgacccagaaatcctatcagaagctgtgtaagaacgtgaagcttggaaat 120
    |||||||
DB 61 caatgtgaacacagaagatatacaagaactgtcaaaaactgaagacttgaagaat 120
QY 121 gccgtcttacccggaataagcctgttcagaagaggaagatgtgtcgaatgaagagcg 180
    |||||||
DB 121 gcagattgacaggttaagttatcttaaaaagaaatgtaataaagaagaaca 180
QY 181 agtgcacgcggttatacaacacgacacccggttctaaagggtctgtcagcgtgtgc 240
    |||||||
DB 181 agtgcacacgctgtatacaacactacactggttcaaaagggtcagttcagtcaggt 240
QY 241 tcaggtggtctgtgtcctctggtggttcgcgtccgcctccgacgacgctgacacagt 300
    |||||||
DB 241 tcaggtggtcagttcagttcaggtgtcaggttcaggttcaggttcaggttcaggt 300
QY 301 ggtcagtgacaaagcggttccggaacagtcgaagaacacacatcctgcacactct 360
    |||||||
DB 301 ggtcagttgctcaggtgttcaggttaattcaagaacgtaacaaatccttcagataattca 360

```

OY	361	agcgattccgaagcgccctctacgccgacctgaagcaccgagtgaatactctcc	420
Dp	361	agtgatccaatgcataactcttaacgtctgatttaaaacacagatgcgaattctg	420
OY	421	actcaaggagcgtgaagatcccaacgctgtgttcgacctactcaatcatgtcgactg	480
Dp	421	actatcaagaagacctcaaatatccctcaacctcttgatttaactcaatcatgttlaact	480
OY	481	tgtatacaatcatgtgcttccaatatccatctgtagtgcaggttlaagaaagatcaatga	540
Dp	481	tgtatataatcatctgttcttcaatatcttaattgatgaggtatgaagaataatgaata	540
OY	541	ctgtacaagttgaattctactctgcactgtgctaagggccaacgtgaatgttgcgc	600
Dp	541	ttatataaatttaacctttattttgtatttaagcgcaaatltaatgatgtatgtcgt	600
OY	601	aatgacattgtcaaatctcatctcaatttgaagatcagagccaagagtttgactatg	660
Dp	601	aatgattatcttccaataaacttccaactcttaaatctcgtgcgaatgaattagactt	660
OY	661	aagaagttgtcttcgcatctcgaagcctctcgacaacatcaaggaacatgttggaa	720
Dp	661	aaaaaactgtgttcgcatatgagaacacattagaacatattaaagatcaatgttga	720
OY	721	atggaagattatattaaagaatlaagaaagaccatcgcgaacatlaacgagctgtat	780
Dp	721	atggaagattacattaaaaaaataaaaaacatagaanaatataatgattatgtga	780
OY	781	gaatccaaaagacccatlagacaaaataagatctgacacagtgatgaagaagaag	840
Dp	781	gaagatlaagaaaacaaatctgataaaaataagaaatgacatlaaagaagaaaaaaa	840
OY	841	tctgacaaagcccgatgcagcgcgtccatctataaacaacagcttgaagaccat	900
Dp	841	ttatccaaagctcaatagtctcttcttattacaataaaacaattgaagaagcacat	900
OY	901	ctcatcagctgactgtgagaagcgcatagacacccctcaagaagaatgaaataca	960
Dp	901	ttataaagcgttttgaanaaacgtatctgcacctttaaanaaaaaatgaaacat	960
OY	961	ctgtctgacaaagttaatgaaatlaagaatctctccgacgaactctgtggacaccc	1020
Dp	961	ttactgtataagaataatgtgaattlaaaaatccccacggtccaattctgtgaat	1020
OY	1021	aacgcgtgtcgtgacaaagaacaaagatataagagcgacgagaagatcaaaaga	1080
Dp	1021	aatatctctctgtataagaaacaaaatacggaggaacgagaagaataaagaat	1080
OY	1081	gccaaaacatlaagtlcaacatagatltctcttcttactgataccctctgtag	1140
Dp	1081	gccaaaacttaaatattaatatgtgtatttattactgataccactggaattga	1140
OY	1141	tacttgagagagaagaataagatatagacatctccgccaagtctgagaagaag	1200
Dp	1141	tatttaagagaaaaaataataaataatctgtataatgtgacaagatgtgaaca	1200
OY	1201	acggaacctaatgaatcccccaatgtgtgagatgacctcttcttaaaagatat	1260
Dp	1201	actgaaacccaatgaaatcccccaatgtgagttacttacccttctgcatataa	1260
OY	1261	aacgtctcaacagatcgaatagcttcggttgactgtgatttaacccctcgatata	1320
Dp	1261	aatgtcttaaatgaacttaattcttctgtgttgaatttaataatccattgtatt	1320
OY	1321	gaacctcttaagaatatctacacagaacaatgtgagaagaagtttatcaacga	1380
Dp	1321	gaacccaagtaaaaacatactactgtataatgaagaagaaaaatctataatga	1380
OY	1381	gagagaatcaaaaattgagaagaagaatgttggagtgagaagaagaagtta	1440
Dp	1381	gaaaaaattaaatgagaaaaaaaaaatttgatctgtataaaaaaactctt	1440
OY	1441	agcaaaagcttlaaacgataatcaatlaaaggatgaaagcgtgtgaacgagat	1500

[illegible]

Db	2521	tctatggtttgtagcttcaacaataggtttatccacaactattatgtaaatatgtaaaaa	2586
Oy	2581	gagatggtctgcaacccgttataaactcaagaacagacagatlaagaacctctggag	2640
Db	2581	gaaatggtttgtaatttatlaaacttaagataatgacaaaattaaattatttagag	2640
Oy	2641	gaagtaagaaggtctccacctctgttaaactctctctccagctccatgaaacctg	2700
Db	2641	gaagcgaaaaaagatattccacactctgttaaaaactcttccaagttcacatgcaacatta	2700
Oy	2701	tctccacaacctcaagaacgaacgggaagtaggtgcttaacgaacacccctcaaccgcgc	2760
Db	2701	tcatataacaccttcgataaaccggaagttagcgcaaatgtagatccatccacctctaca	2760
Oy	2761	aaccttaataactcactgtaaaactgttbgagaacatccctgtctcgcgcagaataagac	2820
Db	2761	aattggaataatagttttaaataattattgtaaaaacactctgactcttggaaaaaaaataat	2820
Oy	2821	attccacaagaactatttgcgacgaataatcgctcgagaacttctcagagaagatctgaa	2880
Db	2821	atatacccaagaattaaataggttccaaaaaagtagtgaaaaactttatgaaaagatlaaaa	2880
Oy	2881	gacacgcgacatctcttaacgaagagcttcaactaaccttgcgttaactctaaagccgaat	2940
Db	2881	gatagtgtacatcttctataatgtaaccttcttacaacttctgttaaatctbaagctgtagt	2940
Oy	2941	atcaacctctcttaacgatalcatctaaacgtlaagaagctggaagagagacatcaatagctg	3000
Db	2941	attaattcatgtgaatgtgtatgcanaaagaagaattagaagaatattataataatta	3000
Oy	3001	aagagaacatcgcacatcgactctgcacctgtacaaagatcaaacatgtaactgtagaaga	3060
Db	3001	aaaaaactcttaccgattatcatctgtattcttataataataataataaattaaattagaaga	3060
Oy	3061	ctcttcgcgaagaagaagacgaactcgcggaagtlaagatgtcgatccaagaagttgactctg	3120
Db	3061	ttatttgttaaaagaagaacacgctgtgttaataataaaycgaaattaaaaaacttactta	3120
Oy	3121	ctcaagagcagcgtctgtaagaacaaactcaactcaactcaactcgcgaaacgttactgag	3180
Db	3121	ttaaagaacatattagatccaaatgtgaattcaacttaataaccacaagaatgtattacaa	3180
Oy	3181	aactcttcagtggtcttccacaagaagaagaagccgagatctcgcgagacagagaacact	3240
Db	3181	aactctctgtttcttcttaacaaaaaaaagaagctgtaaatgacgaagaactgtgaaaaacaa	3240
Oy	3241	ctggagaacacacgaagatcttcttccaacacactacaagaagcctgtcgaattatataatgc	3300
Db	3241	ttagaaaaacaaaaaatatattatgtgaaacattataaagaagctgttaatatataatgagt	3300
Oy	3301	gagcttctcctctgaaagactctctccgaagagagatccacagaccgagataactacgccc	3360
Db	3301	gaatcatcattccattaaanaacttlaagtgaaagaatcaattccaacagaagataatagtcgc	3360
Oy	3361	agcctcgagaacttccaaggtccgtctgaagctcgaagcgaagctgaaggaacaactgaaac	3420
Db	3361	agtttgaanaaacttlaaagattatgaatlaaattagaagaagaaattaaagataattaaat	3420
Oy	3421	ctggagaagaagaagcgcagactccctctcagaagagctgcataccactgtatcgccgaagtc	3480
Db	3421	ttagaaaaaagaanaaatatataactataactatacaagtgagttcacatcattatattgctgaat	3480
Oy	3481	aaggaagatcatlaagaaacaagaactcacacgcggaatagaccgaagcagagataatcacagc	3540
Db	3481	aagaagatataaaaaaataataatatactataccagtgaaattctccagtggaataataacgagt	3540
Oy	3541	gtgataatcagcactgtgaactcttaacaagaagttctcgtctgaaagaaacagatgtccgact	3600
Db	3541	gttaacaatgtattagatcattcaaaaaaattctccgaagagaacagatgtgtcaaca	3600
Oy	3601	gtgtgtctgtaactctgtgctccgcacacacgtgagcagttccaacctaaagaagctgtcact	3660
Db	3601	gtgtlaagtgaaagtgtgattcttcgacacatttagaacaagaagtcacaaagaagaacacgagctca	3660

QY	3661	actcatctgcgagccggtcccaataacatltccacactltccagaacgtgcagatgaggtc	3720
Db	3661	actcatctgcgagccggtcccaataacatltccacactltccagaacgtgcagatgaggtc	3720
QY	3721	gatgcagcatalctgtgcctatctcttcgagagcgagagagactacgatcctgcgc	3780
Db	3721	gatgcagcatalctgtgcctatctcttcgagagcgagagagactacgatcctgcgc	3780
QY	3781	cagctgtgtcaccggtgtgagcgtgtcgaactcctctccgtgtatgtataacatctgtccaaaac	3840
Db	3781	cagctgtgtcaccggtgtgagcgtgtcgaactcctctccgtgtatgtataacatctgtccaaaac	3840
QY	3841	gagaacgatccgagagagcctctatctgaaacccctgcgcggtcctataggtcctcaag	3900
Db	3841	gagaacgatccgagagagcctctatctgaaacccctgcgcggtcctataggtcctcaag	3900
QY	3901	aaacagctgtgagaaatacgtatgacctccaalgtccaacgtgaaagagacattctgcacagc	3960
Db	3901	aaacagctgtgagaaatacgtatgacctccaalgtccaacgtgaaagagacattctgcacagc	3960
QY	3961	cgctttaataagagagaaattccaagaagctcttgagagagagacttgattccctataaa	4020
Db	3961	cgctttaataagagagaaattccaagaagctcttgagagagagacttgattccctataaa	4020
QY	4021	gacctgacctctcctaactcgtgtgtccaagagaccatacaagttccctcaataaagagag	4080
Db	4021	gacctgacctctcctaactcgtgtgtccaagagaccatacaagttccctcaataaagagag	4080
QY	4081	aggagataaattctgccttaattcaactatccaatgagactccatcgacaccgatatacat	4140
Db	4081	aggagataaattctgccttaattcaactatccaatgagactccatcgacaccgatatacat	4140
QY	4141	ctgcctatgtatgtgcgcgggtatcttcaagatccctgagcgaaataatacaagtcgcacct	4200
Db	4141	ctgcctatgtatgtgcgcgggtatcttcaagatccctgagcgaaataatacaagtcgcacct	4200
QY	4201	gacctatataaaagctatccaacgataagcgaagcgagaaatgaaatatctgcacctc	4260
Db	4201	gacctatataaaagctatccaacgataagcgaagcgagaaatgaaatatctgcacctc	4260
QY	4261	ctgataacatccgaaacccgtgtacaagagcgtgaagcgaacaaatcgacctctgcgtatt	4320
Db	4261	ctgataacatccgaaacccgtgtacaagagcgtgaagcgaacaaatcgacctctgcgtatt	4320
QY	4321	caacctgtgagccaaggttcctcaactactactacgaaagagcaatgtggaagttaaaatc	4380
Db	4321	caacctgtgagccaaggttcctcaactactactacgaaagagcaatgtggaagttaaaatc	4380
QY	4381	aaggagctgaaactccccaacaaatccaagaaacgtgcgcagattccaagaagaaataac	4440
Db	4381	aaggagctgaaactccccaacaaatccaagaaacgtgcgcagattccaagaagaaataac	4440
QY	4441	aattctgtcggaaattgtgacagcctgtctaccgatatataccaacaacaatctctgcacag	4500
Db	4441	aattctgtcggaaattgtgacagcctgtctaccgatatataccaacaacaatctctgcacag	4500
QY	4501	ttctgtccactgtgcaggtgttcgaaacccctgcgaacagctgtcgaacatctgctc	4560
Db	4501	ttctgtccactgtgcaggtgttcgaaacccctgcgaacagctgtcgaacatctgctc	4560
QY	4561	gagcgcaacctgcaggtcagctgtgaaacatctcccgacccaatgtgtgaagaaacgtgc	4620
Db	4561	gagcgcaacctgcaggtcagctgtgaaacatctcccgacccaatgtgtgaagaaacgtgc	4620
QY	4621	ccccagataagcggctgttcttaagcagctctgacgcagcgcggaaggtgcaagtgtctcctg	4680
Db	4621	ccccagataagcggctgttcttaagcagctctgacgcagcgcggaaggtgcaagtgtctcctg	4680
QY	4681	aactacaacaagaagagagataaggtgcgtgaggaacccaacacctactctgcaatgaaac	4740
Db	4681	aactacaacaagaagagagataaggtgcgtgaggaacccaacacctactctgcaatgaaac	4740

```

0Y 4741 aatgctggggtgtgagccgcatgtctaahgacccgggaagacagcgctctaagyyaag 4800
Db 4741 aatgctggggtgtgagccgcatgtctaahgacccgggaagacagcgctctaagyyaag 4800
0Y 4801 aaatcacatgtgagtgtaactaacgacactctatccactcttcgaagggatttttcg 4860
Db 4801 aaatcacatgtgagtgtaactaacgacactctatccactcttcgaagggatttttcg 4860
0Y 4861 tccagctcaatttccctgggcatctctctctctgcatcctcatgtcatcctgtacagc 4920
Db 4861 agtctctctaactctttagaataatcatcttataactatcatgttaatatatacagt 4920
0Y 4921 ttcattcataatagatcgatg 4940
Db 4921 ttcattcataatagatcgatg 4940

RESULT 3
AAN50530
ID AAN50530 standard; DNA; 5760 BP.
XX
AC AAN50530;
XX
DT 30-SEP-1991 (first entry)
XX
DE Sequence encoding the P195 protein of Plasmodium falciparum.
XX
KW Malaria vaccine; epitope; antigen; immunogen; ss.
XX
OS Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT CDS 216..5179
FT /tag= a
XX
PM EP154454-A.
XX
PD 11-SEP-1985.
XX
PF 21-FEB-1985; 85EP-0301173.
XX
PR 26-SEP-1984; 84GB-0024340.
PR 22-FEB-1984; 84GB-0004692.
PR 21-FEB-1985; 85GB-0004429.
XX
PA (WELL ) WELLCOME FOUNDATION LTD.
XX
PI Holder A, Sandhu J, Odink K, Lockyer M, Riveros-Moreno V,
DR WPI: 1985-224845/37.
DR P-PSDB: AAP50777.
XX
PT Cloned DNA sequence encoding plasmodium falciparum protein -
PT useful for expressing the protein for use in vaccines against
PT malaria
PT
PS Claim 2; Fig 1; 51pp; English.
XX
CC The sequence encoding the P195 protein of Plasmodium falciparum
CC (AAN50530) and a peptide comprising at least one of its epitopes
CC (see AAP50777) are claimed. Also claimed is a vaccine for inducing
CC immunity to malaria comprising the novel peptide or P195 or a
CC peptide comprising at least one epitope when derived from the new
CC DNA sequence, together with a carrier.
XX
SQ Sequence 5760 BP; 2565 A; 630 C; 725 G; 1840 T; 0 other;

Query Match 54.8%; Score 2706.2; DB 6; Length 5760;
Best Local Similarity 72.1%; Pred. No. 0;
Matches 3587; Conservative 0; Mismatches 1343; Indels 45; Gaps 3;
10 atgaaatattttctctctgtcatcttctgttttataatcaatcaatgagtgctg 69

```

Db	216	atgaagatcatattcttctttagtcaattctcttcttcttattatataataacacatggtga	275
OY	70	accacgaatctcttcagagcgctgtgtttaagaaacttgaaagcttggaaatgcccctct	129
Db	276	acacatgaagttctcaagaacttgcacaaaaactgaagccttctgaagaatgcagttatg	335
OY	130	accggtatacagcgtcttcacgaagagaaatgtgtctgaatgaagggagcagttgacag	189
Db	336	acaggttatagttattcttccaaaagaaaatgttattaatgaagagaaacagtgaaaca	395
OY	190	gccgtttacacacagacacccggttctaaaggctctgtgctagcgtgtgctcogtggg	249
Db	396	gctgttacaacttagtaaaccttggttcaaaaggtctcagttgcttcaggtgtgtcaggtggc	455
OY	250	tctgtgaccttgggggttccgtcgcgtcccgcgaggaggtggcctcaggtgtgcaggt	309
Db	456	tcaagttgtctcaagtggtccagttgcttccagggtgcgtcaggtgcgtccaggtggctcagtt	515
OY	310	gcaagcgcgcgtttccgggaacagttcgaagaccacatccatcgtacaactctagcgtatcc	369
Db	516	gcttcaggtgtgttcaggttaattcaagacgtaacatcccttcagataatctcaagtgattca	575
OY	370	gaacgcgaagttccctcagcgcagcctcaagcacagsgtgaagaactatctccctacataag	429
Db	576	gatgtcaatcttcaactgttgatttaaaacacagatgacgaattactgttacctatcaaa	635
OY	430	gagcgtgaagttcccaacagttgttcgcagccttactacatcatgtctgaactgtgtgatac	489
Db	636	gaacctcaaatctccctcaactcttgatttacttactacatcatatgttgaacttgtgtgataat	695
OY	480	attcaatgcttcaaatcatctcgattgaagcgttcaagagagatcaatgaaactccctgtgacaag	549
Db	696	attcagtttctcaaatatatttaattgtatgtgatagataagaagaatcaatgaattatataaaa	755
OY	550	ttgaattctcaactgcagcttgcctaagggccaaactgaaatgcagtttgcgcgaatgactat	609
Db	756	ttaactcttattcttgattatttaagggcacaattaaatgactatgtatgtcattatgattatt	815
OY	610	tgtcaaatctcatcattcaatttgaagaatcagagccacagatgtgacgttatgaaagaattg	669
Db	816	tgtcaaatctcatcattcaattcttaaatctgtgcgaatgaaatgagcttactttaaanaactt	875
OY	670	gtcttcggaattctgcaagcctcttcgacaacatcaagagacaaatgtgtggaaagtgaagaat	729
Db	876	gtgttcggaatttagaaaaccatagacatatataagaataatgttggaaaaatgtgaagat	935
OY	730	tatatlaaaaaaataaagaagacatctgaaacattacagagctgtatcgaaatattccaa	789
Db	936	tacattaaaaaaaataaaaaaaccatagaanaatatataatgtaatttgaagaagaatgaag	995
OY	790	aagaccatagacaaaataaagaatgtcaaccaagtgagtgaaagaagaagaatgtgtacag	849
Db	996	aaaaaatctgtttaaaaaataagaaatgtcaactcaaaagaaagaaaaaaatlaatacaa	1055
OY	850	gcccgatcgcgcctgtccatctataaaaaacagcttgaagaagcccataaactctacgcg	909
Db	1056	gtccaatatgtcttcttatttcaataaacaatttgaagaagcccatatttaataagc	1115
OY	910	gtactgtgaagcgtcatagacacctcaagaaagaaatgaaatatataaaagaactgtgtcac	969
Db	1116	gttttagaaaaaagtattgtgcactttaaanaaaaaatgaaacatttaagaaattactgtat	1175
OY	970	aagattaatgaaatlaagaatctctcgcgcagcgaactctcgggaaacacccctaacagctg	1029
Db	1176	aagataaattgaatttaaaaaatcccccacgcggccaattctgtgaaatacacccaattactc	1235
OY	1030	ctggacgaagacaagaagtatagagggcagcaggaagaagatacaagagatcgcacaaac	1089
Db	1236	cttgtataagacaaaaaatcggaggaacagaaagaataaaaaagaattgtgccaacct	1295
OY	1090	attaagttcaacatagattctctcttactgatacccttggagcgtgagttactattgga	1149

D	1296	atuaatttaataatgatgaatgttaatttaattactgatlcacattgaaattgaatactatttaaga	1353
Q	1150	gagaagaataagaatalagacatcccgccaaagtcgagacaaagaaatcacaacgaacct	1209
D	1356	gaaaaaataaaaaatattgtataatgtaaaaagtgtaaacaaagaaatcactgaaccc	1415
Q	1210	aatgaatatcccaatgtgtggaagctaacctctgtcttaaaagatatcaaacagctctc	1269
D	1416	aatgaatcccaatgtgagttacttatccctgttcaataacgaatatcaacaatgtctta	1475
Q	1270	aacgagctcgaatagcttcggttgacttgattaaaccccttgattatcagaaagaaacctt	1329
D	1476	aatgaacttaattcttttgtgtgttaataatccacttgattatacaaaaagaccgaat	1535
Q	1330	aagaatatctacacagacaaatgagagaaagaaatltalcaacgaaatcaagagagaagtc	1389
D	1536	aaaacatatatactgtataatgagaaaaaaattcataatgaaatlaagaaaaaaatt	1595
Q	1390	aaattgagagaagaagaanaattgtagatgtgcacagaagaaagttaagaagaccgcgaagaat	1449
D	1596	aaaaatgaaaaaaaataattgtaactgtataaaaaattcttaagaaagacagatctaagctt	1655
Q	1450	ctaacagatlcacataaaggtatgtaaaagctgtgtgaacgagatctatgatccaatctc	1509
D	1656	ttaaatgatataacaaaagaattgtgaaaaaatcctaactgaatattatgatagcaaatctc	1715
Q	1510	aacaatacaltcgactcgactacccaacttcgagaaatgatgtgaaaaacggtactctacaaa	1569
D	1716	aataatataatagtttaacttaattctcgaaaaaaatgatgtgtgtaaaagatatcatataaa	1775
Q	1570	gtgtagaagaattgacacacccataactctgtcatctatgagaaattctaagcataactt	1629
D	1776	gtgtagaacacttaacaccccttaactattgtgcctcctaagaaattcttaacataactt	1835
Q	1630	gagagaagctcacaagaagcttctaagtattgtagggagactatctctcgtggacaattgtgtg	1689
D	1836	gaaaagcttaacaaaagctcctaataatatacgtgaagatatactcttaagaatatatgaatt	1885
Q	1690	gagaaaagaactaaagattatccaagaatctcatagtaagatcgaacacgagatcgagacg	1749
D	1896	gaaaagaatataaataattataaanaatttaataagaanaatagaataatgagattgaaaca	1955
Q	1750	ctgtttgagaacattaaagaagatgtgaagaacagttgtttggaagaaagaaattacaagaac	1809
D	1956	ttagtgtaaataatlaaaaaagaatgaagacagcttcttgtaaaaaaaattatcataagac	2015
Q	1810	gaaaataaacagagatgagaagatcccggtgggtctccgatatgtttaaggtccaaagttag	1869
D	2016	gaaaataaacccaggtgaaaaaaatttaagaagatataccgcaattgttaaaagttaacaattcaa	2075
Q	1870	aaggtgtcctcatalgaaacaagatttgaacacaaagaaagactcaactatcttgaagac	1929
D	2076	aagttttatattatgaacaaaatttgcgcgaatttaaaaagaccctaattgtttaaaaaat	2135
Q	1930	gtgtagttaaacaataataacatgtgtgcggaatagttaaacgaggaagaaataagcagaa	1989
D	2136	gtagaatataaacaataataacatgtgtcccaacttcttaacaacaagaanaataagaagaa	2195
Q	2196	ccctattatttaattgtgtgtgaaaaaagaatttgaataatttaaaagtgttcaatgcttaag	2255
Q	2050	gtcgagagctgtgtcaacgagagaagaagaacatataaacgtgaagagcagtcagatatc	2109
D	2256	gtagaatcattgtataaatagagaanaaaaaaacaataaaaaacacaaagctcaatcgatatt	2315
Q	2110	tccgagccttcacagaagagagataacccggacaggtctacacaaacccggacaaag	2169
D	2316	tcggaacacatcaacgcgaagagaaataacagagacaagcaacttaacaaacctgtgacaaca	2375
Q	2170	gcgcgttcagctctcgaaagcgtatagctgtgcgaagtcacagcacaagacagaagcagca	2229
D	2376	gcaggaatctgtctttagaagagagattcaagtaacagacaaagcacaagaacaaaacagca	2435

OY	2230	cagctccagcgccagcgagcccggtccagaggtctaaagccagtgctccaccacccagct	2289
OY	2230	cagctccagcgccagcgagcccggtccagaggtctaaagccagtgctccaccacccagct	2289
Db	2436	caaccacccggtaccagtagccagtagcaagaagcaaaagccaaagccaccacccagca	24935
OY	2290	cctgtgaataaacaagaccggaaatgctcagaaactgacatacctgtgaagagctcatgag	2349
Db	2496	ccagtaataataaactgtaaaatggttccaaattgagtattctctgtaaaataattatcaaa	25555
OY	2350	ttcctgatacatcctcactcactcgcacaaatatatccctcgctctccaaagcattatgac	2409
Db	2556	ttttaataacttaactatataatgcccataatatttggttctcaactcaactatgac	2615
OY	2410	gagaagattcttaacagtaacagataaacaaagagagagagtagtaaatgcccctgtg	2469
Db	2616	gaaaagatattaacaacatatataattacaagaagagaaagcaaatgaattcatgct	2675
OY	2470	gaccacatgagcccgctgcttaacatlatccagaaacacatcccggtatgatatgctc	2529
Db	2676	gaccatattagacttatggtttaataatacaaaataacataccctgtaattgatctatggt	27355
OY	2530	gatgagccacaca-----attct	2547
Db	2736	gatgagccacacagttcttccaaactattatggtttatgaaaagaattggttat	2795
OY	2548	ctctctcaacggttcatgagagataatgaaagagatgctgcgaactgataaact	2607
Db	2796	gtatttatattacttagtgaaatttatgaaagaatggttctgttaattatataaact	2855
OY	2608	aaagaacaagcagaatataagaacctcttgaggagagcctaagaagttccacactctgt	2667
Db	2856	aagatataatgcaaaatlaaaatatattatagagaaagcgaaaagatgccacactgta	2915
OY	2668	aaaactctctctccacgtccacatgacacacactctccacacccccaagacagccga	2727
Db	2916	aaaactctctccacgtccacatgacacacactctccacacccccaagacagccga	2975
OY	2728	gtgagcgctaaagcagacacctctccactcgaccacctaactaatcactcagaaactggtt	2787
Db	2976	gtaagtgcaaatgtatgatacatcatctacaaatttgaataatagttlaaaattattt	3035
OY	2788	gagagactccgctctctccgacagataaagaacattaccagaacttatgagacagaa	2847
Db	3036	gaaacactattgagctctctgtaaaaaaacaaaataataataccagaatattatggtcca	3095
OY	2848	tcgctcgagaaactctacggaagagatactgaaagacagcgacacatctataacgagac	2907
Db	3096	agtagtgaaactttttagaaagataataaagatagtgatacttataatgaactt	3155
OY	2998	ttcgctactctcgtaaatcttaagccgagtagatcaactcttaacgatgaattcaa	2967
Db	3156	tttcaaaattctgtaaaalcttaaaagctgtagatataattcattgaaatgaaacaa	3215
OY	2968	cgtaagaagctggaagagagacataaagaagcgtgaagaaagacactgcaactgagctgac	3027
Db	3216	aggaagaagaattagagaagatataataataaaanaacttcagttatcatattgat	3275
OY	3028	ctgtacaaacagtacaaactcgtaaactcgsgagagactctcgacagaaagaaagacgtgcg	3087
Db	3276	ttatataataataataataaaattgaaagatattttgtataaaaagaaacacgttggt	3335
OY	3088	aagtataagctgagatcaagaagtgtagctctgcgtcaagagacagcttgaagaacaaact	3147
Db	3336	aaatataataatgcaaatlaaaaaacttacttataataaagaacaaatgaaatcaaaatg	3395
OY	3148	aactcactgacacatccgaaacacgtatcgcgaaacttccgtgttcttcaaaagag	3207
Db	3396	aattcacttaataaccacaaagcagtattaccaaaacttctgttcttcaacaaaa	3455
OY	3208	aagaagaacgagatgcgcgagacagaagaacactctgaggaaacacaagatcttctcaaa	3267
Db	3456	aaagaagctgaaatagatgaagaaactgtaaaacacattagaanaacaaaaatatcttgaa	3515

QY	3268	cactacaagaagccctcgcaagaattataatgycgagctctctccctcgtgaagctctccc	3327
Db	3516	cattataaagagacttgtttaataattataatggtgatactccactaaacaattaaagt	3575
QY	3328	gagggagcattccagaccgcggggaataactcagccagcctcgagaaacttaagttcgtct	3387
Db	3576	gaagatcaaatccaacaagaagaataattatgcaggttttagaanaactttaagttaaag	3635
QY	3388	aagctcgaaagcgaagcgtgaagagacaacctggaacctgtgaagaagaagctcgaacctc	3447
Db	3636	aaattagaagaagaataattaaaggatbaatctaattttagaagaagaaaaaatatcatacta	3695
QY	3448	ctctgcgagctcgtcaactcgtatccgcgcgcgtctcaagaaagtcattaaagaagaactac	3507
Db	3696	tcaagaggttttaccatctatttaattgtcgtgattaaagaagttataaaaaataaaattat	3755
QY	3508	accggcaatagcccaagccggaataataacgaagtgtaataacgcctggaattctaacag	3567
Db	3736	acaggtaatctccaagcgaataataacggaagttaacaatgcatctgaactcttaacaa	3815
QY	3568	aagttccctgcgaaggaacaagatgtccgcacatgltgtctcgtatctgcgtccgcaca	3627
Db	3816	aaattctccagagaagaacaagatgtgttcaacagctgtgaagtgaaagtgatccgcaca	3875
QY	3628	ctggagcagctcacaactaagaagcctgcatactacatgtcgcgagccgagtcacaataca	3667
Db	3876	ttagaacaagaagtcacaacaagaacacagcatcaactcatctgtaggcagagctcaacaa	3935
QY	3688	attccacatctcaagaacgcgcgcagttggtctcgaagacgtatcatattgtgcctatctc	3747
Db	3936	ataacaacatccaanaaigtctgaattgaagtagaagcgtatcatatgactcaactatlt	3995
QY	3748	ggcagagcgagggagagcctcgaattgaccctgcgcgcagtggtgtacccggagagctgact	3807
Db	3996	ggagaaatccgaagaagatattgattgatttagacaagaatgtagtaacggggaagcagtaact	4055
QY	3808	cccttcggtgattgataacatctgtgccaaaatcagaaacgtgagaataaagttgcatactg	3867
Db	4056	aattccgtaattgataacatactcttcaaaatgtgaanaagatagatggttttatatta	4115
QY	3868	aaacctctggaagcgctctatagttctcccaagaaacgctggaataaagttgatgacc	3927
Db	4116	aaacctttagaaggtgttttagaagtttaaaaaacaattgaaaaataacgtatgaca	4175
QY	3928	ttcaatgctcaacgtgaagagcattctcgaaacgcgcctttaaagaagaagaattcaag	3987
Db	4176	tttaatgttaatgttaagatattttaaattcaagatttaataaagcgtgaanaatttcaaa	4225
QY	3988	aagctcttggaagcgactgtattctccataaagaacctgacctcttaactagctgtgc	4047
Db	4236	aatgttttagatcaggttttaattccatttaagaatttaacaatcaagttaatattgtgtc	4295
QY	4048	aaggaacctatacagttctcccatataagaagaagagagataaatttcgtctagtacaac	4107
Db	4296	aagatccataataattctctataaagaaaaaagagataaattcttgaagagttataat	4355
QY	4108	tatctagaagcctcatcgaccacccgataccaatttcgctaagatgtgcgtggattatc	4167
Db	4356	tatattaagatccaatagataagataaataatttgcacaatgagtgctcttggaatatat	4415
QY	4168	aagatcctbgacgnaaaatacaagctcgtgacctgactcattataaaagtataacaagat	4227
Db	4416	aaaaattatctcgnaaaaataataaatacaggttttagattcaattataaaaaatatcaagac	4475
QY	4228	aagcaagcgagagatgaaaaataatctgcacctccctcgtgaataacaatcgaaaacctgtacaag	4287
Db	4476	aaacaaggttgaaattggaataatacctcccttctttaaacaatattggaagactatataaa	4535
QY	4288	acagttgaagcaaaaacgcgcctcttcgtaattcaacctggaagtgccaaggtcctcaacat	4347
Db	4536	acagtttaatgataaaatgtatttatttgaattcaatttgaagcaaaaagttcctaattat	4595
QY	4348	acttacgagaagagcaatgtggaagtttaaattcaagagagctgaactacatcccaacaatc	4407

Db	4596	acatcttggaatcaaacgctgtagaagtttaaaataaagaacttaacttaaaacaactt	4655
Qy	4408	caagacaagcggcgagattccaagaaanaataaactttgcgtggaattgcagacgtct	4467
Db	4656	caagacaacttggcagattttaaaaaaatacaaatlctggttggaattgcgatttaca	4715
Qy	4468	accgatttaaccacaacaactctcttcgacaagtttctcttcacgtgcgcatggttcga	4527
Db	4716	acagttttaaccataataactatttgcacaagttcttccttagtaacgtagtgttttgaa	4775
Qy	4528	aacctgcgaacaacagtgctgcgacatctgcga-cggcaacctgcgaagcgatcgtaa	4586
Db	4776	aactgtctaaatccgltttatcttaacttaattacttgalttggaacttgcgaagtgat	4835
Qy	4587	catctccagacccaactgctgtgaagaacagtg--ccccaagaatagcggtgtttcag	4644
Db	4836	catctcacacaaccaatgctgaanaaaaaaayalccacaagagttctggatgttcaga	4895
Qy	4645	catctgcgcgcgcgcgcgaagtgcaagtgtctctctgaactacaacaagaagagataag	4704
Db	4896	catcttagtgtaagaagaagatgttaattgtttataattacaacaagaagtagtaaa	4955
Qy	4705	tgcgttgagaaccaccaaaccttacttcgaatgaaacaatgycgggtgtgcgcgagtc	4764
Db	4956	tgtgttgtaaaatctcaatcctacttctgaacgaataatgctggtgtgcagatgcc	5015
Qy	4765	aaatgcaccgcggagagacagcgcgtctcaacggaagaanaatacacttgagtgtag	4824
Db	5016	aaatgtacgcgaagaagatctcaggttagcaacggaagaanaatacacttgaagtactaa	5075
Qy	4825	ccgcgactcctacacacttcgcagcggatttttgcgcagctcaattccctggcacc	4884
Db	5076	cctgattgtgtataccacttctgcagtggaatttcttcgagttcctcaactcttgggaata	5135
Qy	4885	tcctctcgtcgtacatccatcgtcgtacgtcgttacagcttcaacttaatagatgat	4939
Db	5136	tcatcttataataacacatcagttataataataacagttcatltaaaaaatgtag	5190
RESULT 4			
AAQ80911			
ID	AAQ80911 standard; cDNA; 5181 BP.		
AC			
AAQ80911:			
XX			
DT	24-AUG-1995 (first entry)		
XX			
DE	Plasmodium falciparum MSA-1 gene CDNA.		
XX			
KM	Plasmodium falciparum MSA-1 gene; recombinant poxvirus;		
KM	multicomponent multistage malarial vaccines; immunogens;		
KM	malaria diagnosis; ss.		
OS	Plasmodium falciparum (p486195).		
XX			
PN	MO9428930-A.		
PD	22-DEC-1994.		
XX			
PF	10-JUN-1994; 94MO-US06652.		
XX			
PR	11-JUN-1993; 93US-0075783.		
PR	09-JUN-1994; 94US-0257073.		
PA	(VIR0-) VIROGENETICS CORP.		
XX			
PI	De Taisne C, Paoletti E, Tine JA;		
XX			
DR	WPI: 1995-036113/05.		
XX			
PT	Recombinant poxvirus contg. Plasmodium DNA in non-essential		
PT	region - useful in vaccines against malaria and for prodn. of		

OY	1771	gatagaagacagctgtgttggaagaagattacaaaagcagcaaatataacccagatgagaag	1830
Db	1861	agtgaaaacaaaattcttagaaaaaaatttlaaaggactaacacattcagatg--ct	1917
OY	1831	atccggaaggtcccggtatgttgaagccagatgcagaagtgctccatgacag	1890
Db	1918	tccttgaagtatatgtatcttgaataataacaaagtcacaaaagtttttatttaaaaaa	1977
OY	1891	atgtatgaactcaagaagactcaactcattctgtgaagaagtgtagttlaaacaaata	1950
Db	1978	atagaagacttaagaagaagatagattatttttlaaaaaatgcacaactaaagaatgatt	2037
OY	1951	catgtccgaatagtataagcagaggaataagcaggaaccatactactccatccgtactc	2010
Db	2038	catltaccaaatattttlaaacaccacaaaataacccagaccattatttattgttatta	2097
OY	2011	aagaaagagatlaagcaaacaggaagtggttcatgcccgaagtcagagagcttatcaag	2070
Db	2098	aaaaaagaagtatgataaatataaaagaatttatccaaaaagtlaaagacagttaa--ga	2155
OY	2071	gagaagaagaacattaaaaacttgaagacagctagataaactccagaccttccacagga	2130
Db	2156	aagacaacagctgttttatacgaattatacacaacctttagtttcagcaagcaaacactg	2215
OY	2131	gagatgaaccggaacaggtcaccaccaagcccggaacaaagggccgtgtccatccgaagc	2190
Db	2216	aagat-gggggttactctcacacacacattatcccatcagaggaacagagaagtaacga	2274
OY	2191	gatgcgtgcgaagctcaagacagaagcgcaagcgagcgcacccctccagtgccagtgcc	2250
Db	2275	gaaacagagaagaagaagaagaacagtagagacaacacacacacaggtaacataacattaca	2334
OY	2251	gttccagaaggtctlaagctccagtgcttcaacacacacagctccgtgtgaataacagcag	2310
Db	2335	ccaaagaagtaaaagttgttgaataattcaatagaaacataagagtaagacaattacaa	2394
OY	2311	aatgtcagcaaacctgactcactcttgaagaagctctatgattcttcgtaatatacctaca	2370
Db	2395	gcttgcgtaaaaaacagtttacttcaagaagaatttagatgaatttttactaaactatata	2454
OY	2371	tgccacaatatatcccgctcctccagcactatagaagaagagaattcttaaacagtac	2430
Db	2455	tgctataatatattttagatataccaaccttagttagagcaaaaaattttagaggtatat	2514
OY	2431	aagaataacaaagagagagaglaaacgctctccctgtgtacacactgagacgtctctc	2490
Db	2515	aattctactccagaagaagaagaatgaataataacatgcatgcatcttagattattattt	2574
OY	2491	aattccagaacaacattcccgtaatagtattctatgttcgatagccccaacattctctc	2550
Db	2575	aatttataaataacatacctctgatagtattctatataatgataagatbaacaatgattta	2634
OY	2551	tctcaactgtctatcagatataatgagaagggagatgctgtaaacctgtataaacctaaa	2610
Db	2635	caacatctcttttctgaattataatcaaaaaggaagatattattatttatacataactaaa	2694
OY	2611	gacaacgacagaattlaagaaacctctcggaagagactlaagaagctc-----	2657
Db	2695	gagggaataatcaatcaaaaaaataattatgaagggcaaaaaaataacatcggaaatcatct	2754
OY	2658	-----caactctgtlaaacatctctctc	2681
Db	2755	acatccagctcttgaaatacaaacgtaaaatactgtcttaaatccgcaactcacagtaattcc	2814
OY	2682	cagctcatgtcaaaccaactgtctctcaactctaagaca-----	2720
Db	2815	caaaaaccaacatcaaatgtatctcttacaatlaaccaaaatgtygtatgctgtatcatct	2874
OY	2721	-----ggccgaagttagcgtctaacgcagacactctcaactgcaccaacttaataactca	2775
Db	2875	ggtctcgtctgtatgtgaagaagatcatgtactcccttaacagatattgtgtatagtataagat	2934
OY	2776	ctgaacatgttttagaacaactcctgtctctccgycagaaglaaagaat--ctaccaagaa	2832

Db	2935	ttgaaaggatattgttaagttctttaaacttcttggaatactaaactaaagtcacataatccatla	2994
Qy	2833	cttatgtgacagaatctcgctccggaactctttagagaagaatctgagagacgcagccca	28922
Db	2995	accattctacacagagatgaggaataattttagaagaatattttaaataaataatgatacc	3054
Qy	2893	ttctatacgaagagcttcactaaactctgtgaaatctcaaaagccgaatgatatacaactctct	29552
Db	3055	tattttaatgtatgatacaaacatcttcgtaaaatcttaattcaaaagttaattacagtttg	3114
Qy	2955	aacgatgaatttaaagctaaagaagctgaggaagacatcaataagctgaaagaagacactg	3012
Db	3115	acc---ggaacacacaaaaaatgcatlaaatgataatgaaatlaaaagatacttta	3171
Qy	3013	caacgagactctgcgcctgtacaaagaatnaaaactgaaacttgagagagactctgcagag	3072
Db	3172	cagttatcattctgtatatataataataataaattaaattgataagattatlaataaag	3231
Qy	3073	aagaagacagtcggaagaatlaaagatgcagatccaagaagttgactctgcacagagcag	31322
Db	3233	aaaaaagacttggccagagcaaaatgccaatttaaaaaacttactttaataaagaacaa	3291
Qy	3133	cttgaagcaaacaccactcaactggaacaatcgcgaacacgatactgcgaactctcagtg	3192
Db	3292	ttagatccaataatggaattcaacttaataaccacataatgatttacaactttctgt	3351
Qy	3193	ttcttcaacaagaagaagagccggaatcgcgagagacagagaacacacttgagagaaccc	3252
Db	3352	ttctttaacaaaaaaaagaagcctgtaaatgacgaaacctgaaacacacttgaaacacca	3411
Qy	3253	aagattctctccaacacactcaaaagccctgtaagatataataatggagagctctctcc	3312
Db	3412	aaatatattatggaacacttatataagaactgttaaatatattataatgggaatcatctcca	3471
Qy	3313	ctggaagactctctccgagagagcattcccgagccagagataactacgcagccctcgagac	3372
Db	3472	ttaaaaactttaaagtgagatataatccaacacagaagaataattatgccaatttgaagaaa	3531
Qy	3373	ttcaaggtccctgttaagctcgaagaacgaactggaagaaacacttgaaacttgagagaag	3432
Db	3532	tttagagtataatgaataaataagatgtaaaaaactcaatgaataattcaattcggagaagaa	3551
Qy	3433	aagctcagctacactctctagcggagatgcatacctgatctgcggagcctcaaggaagtcatt	3492
Db	3592	aaattatcttcttataccaagtgtgattaccagtttaattactgaatttaaagaagtaata	3651
Qy	3493	aagaaacaaagactcacccggccaatagcccacagcgaaataataatacagacgtgaataacga	3552
Db	3652	aaaaataaaaaattatacaggtlaattctccaagtgaaaaataataaagaagttacagaact	3711
Qy	3553	ctggaatcttcaagaaggttcctgcctcggaaggaacagatgctgcgaactggtgtc---	3608
Db	3712	ttaaaatcttaacgaaaatttcttcccaagacaaaagttlaacaacagttgtaactccact	3771
Qy	3609	-----tgaatctgctccgaacacacggtggagcagtc	3641
Db	3772	caaccagatgtaactcatctcatatctatctgtaagaagtgtaagtgagttcagatccaca	3831
Qy	3642	acctaaagag-----cctgcatcactcatalgtcggagccgagttccaataataatacc	3693
Db	3833	aaagaagaaacacataccaacttcagcctctttaaacaagaattacaacaagtagta	3891
Qy	3694	acatctcgaaacgctcgagcagtgagctgcagttgcagtcatactatgtgctatcttcggagag	3753
Db	3892	caattacaacaaatttagtcgaagaagaatgattcccttagttgtaattacccattttggagaa	3951
Qy	3754	agcgaaggaactacgatacgtaaccctcgccaggtgtgtaccgcgggaagctgtcaactcttc	3813
Db	3952	ttcgaaagataatgacgataatttagatacaagtaacactggggaagc---aatactgtc	4008
Qy	3814	gttatgtataacattctgtcccaaatctcgaaacgaatacgaagtgctatctgaaact	3873

Db 4009 acaatgataatcctcctcaggaattgaaaaatgatatgattatatttaaaccct 4068
 Oy 3874 ctggagagcgtctataggctctctcaagaagaacgctggagaataagtgatgacctcaat 3933
 Db 4069 ttacgtggatatacaagaagcttcaaaaaaacaattgaaaaaacaatttctacattatc 4128
 Oy 3934 gtcaacggaaggacatctctgacagcgctttaaagaagaataattccaagctc 3993
 Db 4129 ttaatttgaacgatatcttcaattcaacgtctcttaagaagaacgaataatttcttaagtta 4188
 Oy 3994 ttggagagcagcttgatctcctataaagaacctgacctcctctaactacgttgcagaagac 4053
 Db 4189 ttgagatctgatttaattgaattcaattacatatatctccaatgatatcatatttaagat 4248
 Oy 4054 ccatcaagttctcctaataagaagaaggatatttctgtctagtctacacattatc 4113
 Db 4249 tcaatttaatttgaattctgacacaaaaaacacactttaaagttacaaatata 4308
 Oy 4114 aagagctcatcgacacagatatcaatttcgttaagtgtggtggtatcaagatc 4173
 Db 4309 aaagatcagtagaagaataatgatatatttgcacaggaagagatagttatataaag 4368
 Oy 4174 ctgagcgaataatcaagacttgaccttgactctatataaagaatatatacagataag--- 4230
 Db 4369 gtttagcgaataataagatgatttagaatcaatcaaaaaagttatacaagaagaag 4428
 Oy 4231 -----caa 4233
 Db 4429 gagaaagttcccatcatcaccacaaacacctccgtcacagtaaaaaacagaagaaca 4488
 Oy 4234 ggcgagaaatgaataatctcctctcctgaataaactcgaaacccgtlacagaacagt 4293
 Db 4489 aagaaggaaglaagttctcctccattttaaacaacatttgagacctatacaatactta 4548
 Oy 4234 aacgcgaataatcgacctctctgttaattcaactgagcgcaaggtcctaactaactaac 4353
 Db 4549 gttataaataatgacattacttaacttaagcaagatataagattgattgt 4608
 Oy 4354 gagaagagcaatgtgaggttaaaatcaagaagagctgaactaactaacaatccaagac 4413
 Db 4609 gaataaggtgagccacatggttaaaataaactaactagttgatttaaaagcaattgagac 4668
 Oy 4414 aagctgagagatttcaagaataatatacaatttctgctggaattgacagaccttctacgct 4473
 Db 4669 aaatagatctttaaacaacatacagactcgagacgaattaaataatgataatagat 4728
 Oy 4474 tataaccacaacatctcctgaacgaatttctgtccactgcatggtgttcgaaaacctc 4533
 Db 4729 gatacgaataaagatatactgtgcgaattacttagtacagagtag---ttcaaaattt 4785
 Oy 4534 gccaaaacagtgctgagcaatctgtcgacgcaacctgcaaggagcatgctgaacatctcc 4593
 Db 4786 cctaatacaataatacaataatataatgaagaagaataatccagaataatgttaaacatttca 4845
 Oy 4594 cagcaccaatgctggaagaacacagtgccccagaatagcggctgtttcaaggcatctgac 4653
 Db 4846 caacaacacatgctgaataaaaaacaatgctcagaataattctgtagtttcagacatttagat 4905
 Oy 4654 gagcgcggaaggtgcaatgctcctgaactacacaagaagaaggagataagtgctgag 4713
 Db 4906 gaagaagagaatgtaaatgtttatttaattatacaacaagaagggtgataatgttltgaa 4965
 Oy 4714 aaccaaacctcactctgtaagaataatgacgaatgctgagcgcgatgcttaaatgaccc 4773
 Db 4966 aatccaatactctacttctgaagaataatggtgtgagtgtgacgagacgaataatgaccc 5025
 Oy 4774 gaggaagacagcggtcctaagaagaataatcaatgctgaggtgactaagcccgatcc 4833
 Db 5026 gaagaagatcaggtgagcaagaagaataatcaatgtagatgactaacaacctgattct 5085
 Oy 4834 tatcactcttcgacgagatttctgtccagctcctaatttccctggacatctcctctg 4893
 Db 5086 tatccacttctcagtgatttctgtcagttcctcctaactttagaataatcattctta 5145

Oy 4894 ctgactccatcgtcgtatcgtctacagcttcaatc 4929
 Db 5146 ttaatacctcagtgataataataatacagttcatttaa 5181

RESULT 5

AAT97956
 ID AAT97956 standard; DNA; 1950 BP.

XX AAT97956;

DT 03-APR-1998 (first entry)

DE Chimeric MSA-1 antigenic protein 1 used in a malaria vaccine.

KW Signal peptide; malaria; vaccine; merozoite surface antigen-1 peptide;
 MSA-1; recombinant vaccinia virus; Plasmodium falciparum; anchor peptide;
 immune response; humoral; cell-mediated; merozoite; ss.

OS Chimeric - Mammalia.
 OS Chimeric - Plasmodium falciparum.

OS Chimeric - Homo sapiens.

FH Key Location/Qualifiers
 FT 1..165

FT misc_feature
 FT 166..1893
 FT /tag= a
 FT /note= "signal sequence added to improve immunogenicity"

FT misc_feature
 FT /tag= b
 FT /note= "Plasmodium falciparum MSA-1 peptide"

FT 1894..1950
 FT /tag= c
 FT /note= "anchor sequence added to improve immunogenicity"

PN W09726911-A1.

PD 31-JUL-1997.

XX 29-JAN-1997; 97W0-US01395.

XX 29-JAN-1996; 96US-0593006.

XX (GBOU) UNIV GEORGETOWN.

XX Davidson EA, Yang S;

XX WPI; 1997-393372/36.

PT Malaria vaccine - comprises expression vector expressing fragment of
 PT merozoite surface antigen.

XX Claim 33; Fig 2; 75pp; English.

CC The present sequence encodes a chimeric protein that contains a
 CC mammalian signal and anchor sequence, and a merozoite surface
 CC antigen-1 (MSA-1) carboxy terminal peptide. The C-terminal fragment of
 CC MSA-1 provides a more specific response than the complete MSA-1,
 CC and attachment of anchor and signal sequences improve the immunogenicity
 CC of the protein better than the use of an adjuvant. The chimeric protein,
 CC and expression vectors (analogues that express MSA-1 without either
 CC signal or anchor peptides), particularly in the form of recombinant
 CC vaccinia virus, are used in vaccines to prevent or treat malaria caused
 CC by Plasmodium falciparum. The vaccinia vector expresses the antigen
 CC fragment for many days, or even years, generating a long-lasting immune
 CC response (humoral and/or cell-mediated) against the merozoite form of
 CC the parasite, in humans or other animals.

Sequence 1950 BP; 830 A; 236 C; 277 G; 607 T; 0 other;

Query Match 20.2%; Score 1000; DB 18; Length 1950;
 Best Local Similarity 72.4%; Pred.No.2.6e-233;
 Matches 1297; Conservative 0; Mismatches 495; Indels 0; Gaps 0;

QY 3138 aagcaactcaactctacggaacatccgaacgctactgcgagactctcagttctt 3197
 DB 159 aaatgaattgaattactactaaacccaagatgattacaaaactttctgttctt 218
 QY 3198 caacaagaagaaggagagatcgccgagacagagacatcgccggagaaacccaagt 3257
 DB 219 taacaaaaaaagaagctggaatagcagaacgtaaaacacattagaaaaacaaaaa 278
 QY 3258 tcttcacaacactcaaaagcctcgtaagatataataggcgagcttctccctcgaa 3317
 DB 279 attattgaaacattaaaggactgtttaattataatgttgatcctccctaa 338
 QY 3318 gactctccgagagagacatccgacgagataactaagccgactcgagaactcaa 3377
 DB 339 aacttlaagtgaaagatcaatccaacagaaatattatgccaagtttagaaaactttaa 398
 QY 3378 ggtcctgtctgaagcgaagcaagctgaagacaacctgaagcctggagaaagaagct 3437
 DB 399 agtattaagtaaatagagaagaaatlaagataatlaattagaagaagataaaaa 458
 QY 3438 cagctacctctcagcagactgcatcactgacgcgcgagctcaaggaagtcattaaaga 3497
 DB 459 atcatactatcaagtgagattacatcatcttaatttgctgaattaaagaagtataaaaa 518
 QY 3498 caagaactacacggcaatagccaagcgagataatacagaagctgaatacgcactgga 3557
 DB 519 taaaattatcacagtaattctcccaagtgaaataatcacgtagttaaacaatgcatlaga 578
 QY 3558 attcttaagaagaattccgcctcgagagacagatgtgcacctgtgtgtcgaattcgg 3617
 DB 579 atcttacaacaaatctcccccagaagacagatgtgcacaagctgtgaagaaagtg 638
 QY 3618 ctccgacacactgagcagctctcaacctaaagaagcctcatctactatgtctcgagcgca 3677
 DB 639 atccgacacattagagacaagtcacacaaagaacacgcatcaactatgtctggcgca 698
 QY 3678 gtccaatacaattaccacatctcgaacgctcgagatgaagtgactgaagtcactatgt 3737
 DB 699 gtctacacaaatcaacaacatcacaaatgtcgatgataagtagtcgatacatag 758
 QY 3738 gactatctggcgagagcgagagactacogatacctcgccagcggtgtacccgggca 3797
 DB 759 acctatatttggagatcccgagaagatltatgatttagtagcagaagtacagagaga 818
 QY 3798 ggcgtcacctccttcgagattgatacatctgtccaaatcggaacggaatcacga 3857
 DB 819 agcagtaactccttcgcgaattgataacatacttcttaaatgtgaatgaatgaag 878
 QY 3858 gctctatcgaacactctgacgagcgctctatagctctcaagaacagctlgagagataa 3917
 DB 879 ttatattaaacaccttagcaggtgtttatagaaagttlaaaaaacatttagaataaa 938
 QY 3918 cgtgagtgactctcaatgcaacgctgaagacatcttgaacagcgcttataagaaga 3977
 DB 939 cgtatgacatttaattgttaattgaagatlttaatttaacgatttataaacaagctga 998
 QY 3978 aaattcaagaacgcttgagagagcagcttgattccctataagaagcctggaactctcaa 4037
 DB 999 aaattcaaaaatgttttagaalcgatttaattccatataagaagtttaacctcaagtaa 1058
 QY 4038 ctacgtgtcagaagccatcaagcttccatcaaaagaagaaggaataaattctgtc 4097
 DB 1059 ttatgtgtcagaagatccatataaattcttaataagaagaaagataaattcttaag 1118
 QY 4098 tagttacaacatatacaagagactcatcgacacgcatcaattcgcgttaagatgtct 4157
 DB 1119 cagttataatttataaggtatcaatagatagcagataaatttgcgaatgagtgctt 1178
 QY 4158 ggggtattacaagatccctgagcgaaaaatatacagctgaccttgactctatataaagta 4217

DB 1179 tggattatataaataattatccgaaaaataataacagattagattcaattaaataa 1238
 QY 4218 tatcaagatagaagcagcgagaaatgaaaaataatctgccttcccgataaatctgaac 4277
 DB 1239 tatcaagcacaacaaggltgaaaaatgagaaatataccttccctttaaacaatattgagac 1298
 QY 4278 cctgtacaagaagtgaaacgacaacatcgacctctgttaacttcaaccggagcgcaag 4337
 DB 1299 ctatataaacaagttatgataaaatttgatttatttgaattatcattagaagcaag 1358
 QY 4338 cctcaactatacttcaagaagagcaatgtggaagtlaaatcaagagagcgtgaactact 4397
 DB 1359 tctaattatatacatgagaaatcaaacgtagaagttlaaatlaaagaacttaattact 1418
 QY 4398 caaaacatccagaacgaagctgagcagatttcaagaaaaataacaattcgtcogaattgc 4457
 DB 1419 aaaaacaaltcaagaacaattgycagatttlaaaaaataacaattcgttgyattgc 1478
 QY 4458 agacctgtctacgattataaccacaacatctcctggaagtttctgtccactgcat 4517
 DB 1479 tgattatcaacagattataacataactatltgacaaggttcccttagtaacagat 1538
 QY 4518 ggtgtcgaaaacctcgcaaaacagtgctgagcaatcgtctcgacggcgaacctgag 4577
 DB 1539 ggttttgaaaatctgtctaaacccgtttatctatctacttgtagaacttgcaag 1598
 QY 4578 catgctgaacatctccagcaccaatgcgttagaagaacagatgccccagaatagcgctg 4637
 DB 1599 tatgttaaacatttcaacaaccaatgcglaaaaaacaatgtccacaaattctgag 1658
 QY 4638 ttctgagatctgagcagcgagagtgcaagtgctcccgactcaacaagaag 4697
 DB 1659 ttccgacatttagatgagaagagaaatgtaaaagttatlaataatcaacaagaag 1718
 QY 4698 agataagtcgtggaagcccaacccctlaactgcaatgaaacaatltgvcgggtgagc 4757
 DB 1719 tgataaagtggttgaaaatccaaatccactactgtgaagaaataatgttgyatgagc 1778
 QY 4758 cgaatgtaatgacagaggaagacagcgctcaacggaagaaatccatgcagatg 4817
 DB 1779 agatgcacaattgacgaagagatcagtagcaacggaagaaatccatcagatg 1838
 QY 4818 tactaagccgactctctccacttctgacggagtttttggctccactcaatttct 4877
 DB 1839 tactaaactgtatttcaacacttctcgatgtgatttcttgcagcttccctcaactct 1898
 QY 4878 gggcatcctcctgcgtgatcctatgtcgtgactgctgacaacgtcactaa 4929
 DB 1899 aggaatcatcttataataactcaatgttaatatataacagttcattaa 1950

RESULT 6
 AAT97958
 ID AAT97958 standard; DNA: 1897 BP.
 AC AAT97958;
 DE 03-APR-1998 (first entry)
 XX Chimeric MSA-1 antigenic protein 3 used in a malaria vaccine.
 XX
 DE
 XX
 KW Signal peptide; malaria; vaccine; merozoite surface antigen-1 peptide;
 KW MSA-1; recombinant vaccinia virus; Plasmodium falciparum; anchor peptide;
 KW immune response; humoral; cell-mediated; merozoite; ss.
 OS Chimeric - Plasmodium falciparum.
 XX
 XX Chimeric - Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT misc_feature 1..1839
 FT /tag= a
 FT /note= "Plasmodium falciparum MSA-1 peptide"

```

FT      misc_feature      1840..1897
FT      /*tag="b
FT      /note="anchor sequence added to improve
FT      Immunogenicity"
XX      MO9726911-A1.
XX      31-JUL-1997.
XX      29-JAN-1997; 97WO-US01395.
XX      29-JAN-1997; 96US-0593006.
XX      29-JAN-1996; 96US-0593006.
XX      (GEOU ) UNIV GEORGETOWN.
XX      Davidson EA, Yang S;
XX      WPI; 1997-393372/36.
XX      Malaria vaccine - comprises expression vector expressing fragment of
PT      merozoite surface antigen
XX      Claim 33; Fig 4; 75pp; English.
XX      The present sequence encodes a chimeric protein that contains a
CC      human anchor sequence, and a merozoite surface antigen-1 (MSA-1) carboxy
CC      terminal peptide. The C-terminal fragment of MSA-1 provides a more
CC      specific response than the complete MSA-1, and attachment of the
CC      anchor sequence improve the immunogenicity of the protein better than
CC      the use of an adjuvant. The chimeric protein, and expression vectors
CC      (analogues that express MSA-1 without either signal or anchor peptides),
CC      particularly in the form of recombinant vaccinia virus, are used in
CC      vaccines to prevent or treat malaria caused by Plasmodium falciparum. The
CC      vaccinia vector expresses the antigen fragment for many days, or even
CC      years, generating a long-lasting immune response (humoral and/or
CC      cell-mediated) against the merozoite form of the parasite, in humans or
CC      other animals.
XX      Sequence 1897 BP; 815 A; 230 C; 274 G; 578 T; 0 other:
SO

Query Match      20.0%; Score 989; DB 18; Length 1897;
Best Local Similarity 72.3%; Pred. No. 1.2e-230;
Matches 1297; Conservative 0; Mismatches 495; Indels 1; Gaps 1;

```

```

OY      3558 attctacaagaagttcctgcctgaaagacagatgctgcacatgctggtctgaatctcg 3617
OY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      attctacaaaaaattctccccaagaagacagatgctgcacagttgtaagtgaagtcg 584
OY      3618 ctccgacacactggaagcagttcacaactaagaagcctgcatactactatgctcgaagccga 3677
OY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      atccgacacattgaagacaagaagtcacacccaagaagacagacatcaactatgtagcgagca 644
OY      3678 gtccaatcaattacacaatctcgaacgtgcagatgaagtcgtagtcgcatcatgtt 3737
OY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      gtctaacacaataaacaacatcaacaatgctgagtgagtgagtgagtgagtgagtgagtgagtg 704
OY      3738 gctatcttcgagagcgagagagagacagatgacatgacctgcgcagaggtgcacacgtgca 3797
OY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      acctatattggaagatccgaaagagatgtagatgattgagacagtagtaacagcgaga 764
OY      3798 ggtgtcacctccttcgctgattgataacattcgtccaaatcgagaacgataacgaagt 3857
OY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      agcagtaactcctccgtaattgataacattctctaaatltgaatgtagatagatgaggt 824
OY      3858 gctcatctgaaacctgcgcagcgctctatagtgctctcgaagaacgctggagataa 3917
OY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      ttatattaaaccttgaagcaggtgttatagaagttttaaataaacaattagaataa 884
OY      3918 cgtgtagaccttaactgtaacagtgtaagacatctcgaacgcgctttaaagaaga 3977
OY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      cgtatgacatttaagtgttaagttaagaagatatttaattcagcatttaataacgga 944
OY      3978 aaatttcaagaagcgtcttggaagcagacttgattcctataaagaactgcactctaa 4037
OY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      aaatttcaaaaatgltttagaatcagatttaattcatalaagaatttaacatcaagtaa 1004
OY      4038 ctacgttccaagggccatcaaatcgaatctcctaataaagaagaagaggttaattctgctc 4097
OY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      ttatgtgtcacaagctccatataaaattccttaataaagaaaaaaagaatcttgaag 1064
OY      4098 tagttacaactatatacagaagcctccatcgacacgagatcaattcgttaagtgatgct 4157
OY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      cagttataattatataatgaagttcaatagatagatagatagaataaatttgcaaatgtgtct 1124
OY      4158 ggggtattacaagaatcctgagcgaagaaataacaagtcgtgaactgtgctctataaagta 4217
OY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      tggatattataaataatataatccgaaaaaataaataacagatttagatlttaataaataa 1184
OY      4218 tatcaagataaagcgaagcgaagaaatgaaataatctgccttccttcgaataacatcgaaac 4277
OY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      tatcaagcaacaacaaggtgaaatgaaataacatccctcccttttaacaatatgtgagac 1244
OY      4278 cctgtacaagaagtgaaacgcaaaaatcgacactctgtaattcacctcggagggccaagt 4337
OY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      ctataataaacaagttatgataaataattgtatttattgttaattcattagaagcaaaagt 1304
OY      4338 cctcaactatacttaacgagaagcgaatggtgaagttaataacgaagcgtggaactacct 4397
OY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      tctaattatatacatatgagaatcaacagtagaagtttaataaagaagacttaattacct 1364
OY      4398 caaaacattccaaagcaagcgtgcagatttcaagaanaatacaaatctgcgcgaagt 4457
OY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      aaaaaacattccaagaacaattgagcagattttaaaaaataaacaattcgttggaaatgc 1424
OY      4458 agacctgtacccgatttaacacaacaatctcctgcgaagtttcgtccacgtcgcat 4517
OY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      tgatttatcaacagatttaacacaataactatcttgcaaaagttccttagtagcgat 1484
OY      4518 ggtgttcgaaaaacctgcgcaaaacagtgctgagcaatctgctcgacggaactgcaagg 4577
OY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      ggttttgtaaaactctgtctaaaacggtttatcttaattacttgtagtaaaacttgcaagg 1544
OY      4578 catgctgaacatctcccgacccaatgctgtgaagaacaagtgccccgaatagcgctgct 4637
OY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      tatgttaagaacttcaacacccaatgctgttaaaaaaacaatgcccacaaaattctgagtg 1604
OY      4638 ttcaagcatctcgacgagcggaagagtgcaagtgctcctcgaaactacaacaagaag 4697

```

Db 1605 tttagagatttaagtgaagaagaagtgaatgttatttaattacaacaagaag 1664
OY 4698 agataagtgctgagagaccacacactctgaatgaagaatgctgggtgagc 4757
Db 1665 tgaataatgtgtgaataatccactctgttaacgaataatgtgtgagtcg 1724
OY 4758 cgaatctaataatgcagagagacagcgcttaacggaagaagaatacagtcgagtg 4817
Db 1725 agatgcacaaatgtacacgaagaagatcagtgagcaacggaagaatacagtcgagtg 1784
OY 4818 tactaagccgactcctactcactcctcgcagcgagtttttgcctcagcttaattctc 4877
Db 1785 tactaactcattctatccactcttcgactgtgattttctgcagctcctactactctt 1844
OY 4878 gggactcctcctcctgct-gatcctcaatgcagtcagctgttaagcttactca 4929
Db 1845 aggaataatcattctatctaataactactatgttaataatatacagttcattaa 1897

RESULT 7

AAT97957 standard; DNA: 1896 BP.

AAT97957:

03-APR-1998 (first entry)

Chimeric MSA-1 antigenic protein 2 used in a malaria vaccine.

Signal peptide; malaria; vaccine; merozoite surface antigen-1 peptide;

KM MSA-1; recombinant vaccinia virus; Plasmodium falciparum; anchor peptide;

KW immune response; humoral; cell-mediated; merozoite; ss.

OS Chimeric - Mammalia.

OS Chimeric - Plasmodium falciparum.

FH Key Location/Qualifiers

FT misc_feature 1..165

FT /tag- a

FT /note- "Signal sequence added to improve immunogenicity"

FT misc_feature 166..1896

FT /tag- b

FT /note- "Plasmodium falciparum MSA-1 peptide"

XX MO9726911-A1.

XX 31-JUL-1997.

XX 29-JAN-1997; 97WO-US01395.

XX 29-JAN-1996; 96US-0593006.

XX (GEOU) UNIV GEORGETOWN.

XX Davidson EA, Yang S;

XX WPI; 1997-393372/36.

XX Malaria vaccine.- comprises expression vector expressing fragment of

XX merozoite surface antigen

XX Claim 33: Fig 3: 75pp: English.

CC vaccines to prevent or treat malaria caused by Plasmodium falciparum. The
CC vaccinia vector expresses the antigen fragment for many days, or even
CC years, generating a long-lasting immune response (humoral and/or
CC cell-mediated) against the merozoite form of the parasite, in humans or
CC other animals.
XX
SO Sequence 1896 BP; 813 A; 230 C; 272 G; 581 T; 0 other;

Query Match 19.6%; Score 969.6; DB 18; Length 1896;
Best Local Similarity 72.4%; Pred. NO. 6.5e-226;
Matches 1257; Conservative 0; Mismatches 479; Indels 0; Gaps 0;

OY 3138 aagcaaatcaactacacagcaacatccgaacacagctactgagagactctcagttctt 3197
Db 159 aaatgaattgaattcaactaataacacgaagatgattacaacatttctgtttctt 218
OY 3198 caacaagaagaagaagcgcgagatcgccgagacagagacactctgagaaacacaaagt 3257
Db 219 taacaaaaaaagaagcgtgaatagcagaacctgaaacacattagaacacaaaaat 278
OY 3258 tctctcaaacactacaagaagcctcgtaagtattataatgagagcttctctctgaa 3317
Db 279 attattgaaacattataaagaagctgttaataattataatgagatcatctcaattaa 338
OY 3318 gactctccgagagagacatccagacagcagagatactacgcagcctcgcgaacttca 3377
Db 339 aacttgaagtgaagaataaaltcaacagaagaataaltcagtttagaactttaa 398
OY 3378 gtcctgtctaagctcgaaggaagcgtgaagaacacactgagacactgaggaagaagct 3437
Db 399 agttaagtaataattagaagaataataaagataatttaattagaagaagaataatt 458
OY 3438 cagctacactctcgtcgcgagatcagcagatgcgcagagtcgaagaagtcattaa 3497
Db 459 atcattacttaacagtgattacatcattatgtcgtgaataaagaagtaataaaaa 518
OY 3498 caagaacacacccgcaatagcccaagcagagataataacagcagtgtaacgcagctga 3557
Db 519 taaaattatacaggtlaattctccaagtgaaataatacgaatgttaacaatgacttga 578
OY 3558 atctacaagaagtcctcgtcctgtagaagacagatgtgcacactgtgtgtctgatatc 3617
Db 579 attctacaacaaatcttccccaagaagacagatgtgtgacacagttgtaagtgaagtg 638
OY 3618 ctccgacacactgagagcgttcaactaagaagcctgacatctcaatcgtcgcgagccga 3677
Db 639 atccgacacattagacaagaatgacaacaaagaacacagcatcaatcgtgagagcaga 698
OY 3678 gtccaataaattacacatctcgaagcgtgacagatgaggtgagtgatcagtcattct 3737
Db 699 gtctaacaataaacaacatacacaagaatgtcagatgagatgagtcgtaacatagt 758
OY 3738 gctcatctcgcgagagcagagagacatgacatgacacgagtcgcgagtgacagctgta 3797
Db 759 accataatttgagaatcccgagagagatattagatgatttagaacaagtagtaacaagaga 818
OY 3798 ggcgtgcaactcctcgcgtgattgataacatctgtccaaaatcgagaacgaatacgaagt 3857
Db 819 agcagtaactccttcgcgtaattgataacatcttcttaaatgtaaatgataatagtg 878
OY 3858 gctcatctggaacctcgcgagcgtcattagctcctcaagaacagcgcgagagataa 3917
Db 879 ttatatttaaacctttagcaggtgtttatagaagtttaaaaaacattagaataaa 938
OY 3918 cgtgatgacttcaatgtcaacgctgaaggaacattctgaacgcgcgtttaaagaaga 3977
Db 939 cgttatgactttaaattgttaattgttaagagattttaaattcaagattttaaagaagtg 998
OY 3978 aaatttcaagaagctgtgagagcagctgattcctataaagaactgactcctctaa 4037
Db 999 aaattcaaaaatgttttagatcagatlttaattccatataaagaatttaacatcaagtaa 1058

Db 421 atcaacgataagcaggagagacgagaagtaacgtcccttccttcgtaacaacatcagacc 480
Oy 4279 ctgtacagaagcagtgagacaaaatcgacctcttcgtatcccttgaggcgaagtc 4338
Db 481 ctgtacagaagcagtgagacaaaatcgacctcttcgtatcccttgaggcgaagtc 540
Oy 4339 ctcaactactactagagaagaagcaatgtggaagttaaatcaagaagctgacactc 4398
Db 541 ctgaactacacatagtgaagaagcaagctgagagtcgaatcaagaagctgacactc 600
Oy 4399 aaacaatccagaagcagctgagacatctcaagaanaatacaacttcgtcgaaatgca 4458
Db 601 aagaccatccagagtaagctgagcagatctcaagaanaatacaacttcgtcgaaatgca 660
Oy 4459 gacctgtaccgattataacacacacacatccctgacacagttctctgacactgcatg 4518
Db 661 gacctgtaccgattataacacacacacacatccctgacacagttctctgacacgcatg 720
Oy 4519 gttctgaaaacctgcacaaaacagctgtagcaatctgctcgagcgaacctgcaagtc 4578
Db 721 gttctgaaaacctgcacaaaacagctgtagcaatctgctcgagcgaacctgcaagtc 780
Oy 4579 atgctgaacatctccagcaccacatgctgtaagaacagctgcccagaataagcgctgt 4638
Db 781 atgctgaacatctccagcaccacatgctgtaagaacagctgcccagaataagcgctgt 840
Oy 4639 tttagcagctggaagcagcggaagagtgcaagtgctcttcgtaactatacaaaaaga 4698
Db 841 tttagcagctggaagcagcggaagagtgcaagtgctcttcgtaactatacaaaaaga 900
Oy 4699 gataagtcgctggaagcagcggaagagtgcaagtgctcttcgtaactatacaaaaaga 4758
Db 901 gataagtcgctggaagcagcggaagagtgcaagtgctcttcgtaactatacaaaaaga 960
Oy 4759 gatgctaaatgcacagcaggaagacagcggtctcaacggaagaatacaatcagatgct 4818
Db 961 gatgctaaatgcacagcaggaagacagcggtctcaacggaagaatacaatcagatgct 1020
Oy 4819 actaagccgactcctacacactccttcgacggaattttg 4859
Db 1021 actaagccgactcctacacactccttcgacggaattttg 1061

RESULT 9
AA25586
ID AA25586 standard; CDNA: 1065 BP.
AC AAY25586;
XX 02-AUG-1999 (first entry)
DT
DE Merozoite surface protein MSP-1-42 modified cDNA.
XX
KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal; ss.
XX
OS Plasmodium falciparum.
OS Synthetic.
XX
PN M09920766-A2.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22225.
XX
PR 15-MAY-1998; 98US-0085649.
PR 20-OCT-1997; 97US-0062592.
XX
PA (GENZ) GENZYME TRANSGENICS CORP.
PI Chen LH, Meade H;

DR WPI: 1999-302742/25.
DR P-PSDB: AAY05832.
XX
PT New modified recombinant nucleic acid sequences useful for producing
PT malarial DNA vaccine
XX
PS Claim 12; Fig 1; 43pp; English.
XX
CC This novel, modified nucleic acid encodes a 42 kDa C-terminal
CC part (see AAY05832) of malarial merozoite surface protein MSP-1
CC (MSP-1-42), an important target for the development of a vaccine
CC against Plasmodium falciparum. The nucleic acid sequence has been
CC modified compared to the native sequence of MSP-1-42 (see AAY25587)
CC such that 306 nucleotide positions have been replaced to lower the
CC AT content (from 76 to 49.7%) and to eliminate 10 mRNA instability
CC motifs while maintaining the same protein amino acid sequence.
CC These alterations allow MSP-1-42 to be expressed in mammalian cell
CC culture and in transgenic mice. The invention provides modified
CC recombinant nucleic acid sequences and methods for increasing the
CC mRNA levels and protein expression of proteins that are difficult
CC to express in cell culture systems, mammalian cell culture systems
CC or in transgenic animals. The preferred difficult protein
CC candidates for expression are those derived from lower organisms
CC such as parasites, bacteria and viruses that have DNA coding
CC sequences of high AT content or which have mRNA instability motifs
CC or rare codons relative to the recombinant expression system to be
CC used. The invention allows expression of MSP-1 protein in the milk
CC of transgenic animals, and also provides a DNA vaccine comprising a
CC vector containing the altered MSP-1-42 sequence.
XX
SQ Sequence 1065 BP; 324 A; 254 C; 280 G; 207 T; 0 other;

Query Match 15.1%; Score 744.2; DB 20; Length 1065;
Best Local Similarity 81.3%; Pred. No. 3.3e-171;
Matches 863; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

Oy 3799 gctgcatccttcctgtagatataacattctgccaatcgagaagcgaagtg 3858
Db 1 gctgcatccttcctgtagatataacattctgccaatcgagaagcgaagtg 60
Oy 3859 ctctactgaacactctgcaagcgctctatagctctcaagaagaagctggaataac 3918
Db 61 ctgtactgaagcgcgtgcaagcgctctatagctctcaagaagaagcgtggaataac 120
Oy 3919 gtgatgaccttaatgtaacgtggaagacatctgcaacgctgcttaataagaagaa 3978
Db 121 gtgatgaccttaatgtaacgtggaagacatctgcaacgctgcttaataagaagaa 180
Oy 3979 aattcaagaagctgtgagaggaactgtatcccttaagaagcctcccttaac 4038
Db 181 aattcaagaagctgtgagaggaactgtatcccttaagaagcctcccttaac 240
Oy 4039 tacgtgtcaagagaccatacaagttctcctaataagaagaagataaattctgct 4098
Db 241 tacgtgtcaagagaccatacaagttctcctaataagaagaagataaattctgctg 300
Oy 4099 agttacaactatataagaagctcgaacacgataatcgaatctgcaatgagtgctg 4158
Db 301 agttacaactatataagaagctcgaacacgataatcgaatctgcaatgagtgctg 360
Oy 4159 gggtattcaagaagctgagcgaataatcaagctgagccttgatctataaagaat 4218
Db 361 gggtattcaagaagctgagcgaataatcaagctgagccttgatctataaagaat 420
Oy 4219 atcaacgataagcaggagagatgaataatactgaccttccttcgtaataacatcgaaac 4278
Db 421 atcaacgataagcaggagagagatgaataatactgaccttccttcgtaataacatcgaaac 480
Oy 4279 ctgtacagaagcagtgagacaaaatcgacctcttcgtatcccttgaggcgaagtc 4338
Db 481 ctgtacagaagcagtgagacaaaatcgacctcttcgtatcccttgaggcgaagtc 540

```

QY 4339 ctcaactactacagagaagcaatgtggaagttaaatcaagaagctgaactacc 4398
    || ||||| || || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 ctgactactacatagtgaagaagcaacgtgaggtcaagatcaagaagctgaattactgtg 600
QY 4339 aaacaataccaaagaagctggtcgaattctcaagaataatacaattcgttcgaattcca 4458
    || || ||||| || || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 601 aagaccatccaggaataagctgtccgatttcaagaagaacaacaattcgttcggaatccg 660
QY 4459 gacctgtaccagattatataacacaacatctcctgaccagatttcttccactgtgcatg 4518
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 661 gatctgagacacgattataacacaacaacactcgtcgaagaattccttgaacacggtatgt 720
QY 4519 gtgttcgaaaacctgcgcaaaaacagtgctgagcaatctgctcgaagcgaacctgcaagggc 4578
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 gtcttcgaaaacctgtgcaagaacgctcctgagcaacctcgtgtgagtggaactcgaagg 780
QY 4579 atgtctgaacatctccagacacacatgtgtgaagaacaagtgccccagaatagcgctgt 4638
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 atgtcgaacatctcagcagacacacagtgltggaagaagcagtgctcccgagaacagcggtgt 840
QY 4639 tttaggcacatctgagcagcgcggaagaagtgcaagtgctcctcgaactatacaacaagaaga 4698
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 841 tttaggcacacccgagagagagagagagtgtaagtgctcgtgaactacaagcagaagaagtc 900
QY 4699 gataagtgctgagagaaacccaacctactcctcgaatgaagaacaatgagcggtgtgacgcc 4758
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 901 gataagtgctgagagaaacccaacctactcctcgaatgaagaacaatgagcggtgtgacgcc 960
QY 4759 gatctcaaatcaccagagagagagacagcgctcctcaaggaagaataacatcagtgcgatgt 4818
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 961 gatcccaagtgctacagagagagatcagagagagcagcaggaagaagatcacctgtgagtg 1020
QY 4819 actaagcccgactctcactcactcctcctgacgagattttt 4859
    || ||||| || || ||||| ||||| ||||| ||||| ||||| |||||
Db 1021 accaagcctgattcttactcactgttcgattgtactctctg 1061

```

RESULT 10 AA56021

ID AA56021 standard; CDNA; 1140 BP.

AC AA56021;

DT 31-AUG-1999 (first entry)

DE Merozoite surface protein MSP-1-42 modified cDNA.

KW MSP-1: merozoite surface protein; malaria; vaccine;

KW protein engineering; protein expression; codon usage;

KW transgenic animal; mutant; ss.

OS Plasmodium falciparum.

XX Synthetic.

XX Key Location/Qualifiers

XX CDS 1..1131

XX sig_peptide 1..45

XX mat_peptide 46..1128

XX WO9920774-A2.

XX PD 29-APR-1999.

XX PF 20-OCT-1998; 98MO-US22226.

XX PR 15-MAY-1998; 98US-0085649.

XX PA 20-OCT-1997; 97US-0062592.

XX (GENZ) GENZYME TRANSGENICS CORP.

```

PI Chen LH, Meade H;
XX WPI: 1999-288313/24.
DR P-PSDB: AA09374.
XX Modified malarial protein for use in anti-malarial vaccines
XX PS Example: Fig 11; 35pp; English.
XX This novel, modified nucleic acid encodes a 42 kDa C-terminal
CC part (see AA09374) of malarial merozoite surface protein MSP-1
CC (MSP-1-42), an important target for the development of a vaccine
CC against Plasmodium falciparum. The nucleic acid sequence has been
CC modified compared to the native sequence of MSP-1-42 (see AA56009)
CC such that the AT content has been reduced and 10 mRNA instability
CC motifs eliminated while maintaining the protein amino acid sequence.
CC In addition, a sequence encoding a 15-amino acid beta-casein signal
CC peptide has been added to the 5' end of the sequence, and N262Q
CC and N181Q mutations have been introduced to eliminate
CC N-glycosylation sites. These alterations allow MSP-1-42 to be
CC expressed in the mammary gland (i.e. milk) of transgenic mice. The
CC invention also provides a DNA vaccine comprising a vector containing
CC an altered MSP-1-42 sequence.
XX S0

```

Sequence 1140 BP; 353 A; 282 C; 290 G; 215 T; 0 other:

Query Match 15.0%; Score 742; DB 20; Length 1140;
Best Local Similarity 80.8%; Pred. No. 1.2e-170;
Matches 865; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

```

QY 3791 ccggtgagcgtgcactcctccgtgattgataacattcttccaaatcgaagaacgat 3850
    || || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38 ccatgcagccgtcactcctccgtcattcgaataacattcttccaaatcgaagaacgat 97
QY 3851 accgaagtgctctatcgaacacctctgcagcgctctataggtctctcaagaacagctgtg 3910
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98 accgaagtgctctatcgaacacctctgcagcgctctataggtctctcaagaacagctgtg 157
QY 3911 agataaagtgatgactcctcaatgtcaacgtggaagaatctcgaacgcccgttata 3970
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 158 agataaagtgatgactcctcaatgtcaacgtggaagaatctcgaacgcccgttata 217
QY 3971 agagaagaattccaagaacgtcttggaagcgaactgtatccctataaagaactgact 4030
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 218 agagggagaactccaagaacgtcttggaagcgaactgtatccctataaagaactgact 277
QY 4031 cctcaactagtgltcaagaacccaatacaagttcctcaataaagaagaagataat 4090
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 278 gaagaactacgtgtcaagaacccaatacaagttcctcaataaagaagaagataat 337
QY 4091 ttctgtcagttacaacatataccaaggtcctcgaacccgatacaatctgtctaatg 4150
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 338 ttctgtcagttacaacatataccaaggtcctcgaacccgatacaatctgtctaatg 397
QY 4151 atgtcgtgggtattacaagaactcctgagcgaagaataatacaagttcgaacttata 4210
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 398 atgtcgtgggtattacaagaactcctgagcgaagaataatacaagttcgaacttata 457
QY 4211 aaaagtatatacaagataagaagcgaagaatgaaatatacttccctctcgaataaca 4270
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 458 aagaatatacaagataagaagcgaagaagtaagttaccttccctcgaataaca 517
QY 4271 tcgaaacctgtacagaacatgaaacgaataatcgactcttgtaattcaactcgtgagg 4330
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 518 tcgaaacctgtacagaacatgaaacgaataatcgactcttgtaattcaactcgtgagg 577
QY 4331 coaaggtcctcaactatactacgaagaagcaatgtggaagttaaaatcaagaagctga 4390
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 578 coaaggtcctcaactatactacgaagaagcaatgtggaagttaaaatcaagaagctga 637
QY 4391 actactcaaaacaatccaagaacagctgtgcagatttcaagaagaatacaattctgtcg 4450
    || ||||| || || ||||| || ||||| || ||||| || ||||| || |||||

```

Dh 638 attactgaagaccatccagataaagctgctgcatlctcaagaagacacacttcgtcg 697
Qy 4451 gaattgcagactctctaccgattataaccacaactctctgaccagttctgtcca 4510
Dh 698 gaatcgccgactctgagaccgattacacacacaaactctgacccaagttcttgaca 757
Qy 4511 ctgcagatggtcttcgaanaaccctgcgaanaacagtcgtgacactctgacggaacc 4570
Dh 758 ccggaatggtcttcgaanaaccctgcgaanaacagtcgtgacactctgacggaacc 817
Qy 4571 tgcagggcatctgacactctccagacacaaatgctgtaagaanaacagtcgtccagaata 4630
Dh 818 tgcagggatctctgacactctgacacacaaatgctgtaagaanaacagtcgtccagaata 877
Qy 4631 gcgcgtcttctgacactctgacacacaaatgctgtaagaanaacagtcgtccagaata 4690
Dh 878 gcgcgtcttctgacactctgacacacaaatgctgtaagaanaacagtcgtccagaata 937
Qy 4691 aagaagagataaagtcgtgtaagaacacaaactctgacacacaaatgctgtaagaanaac 4750
Dh 938 aagaagagataaagtcgtgtaagaacacaaactctgacacacaaatgctgtaagaanaac 997
Qy 4751 gtcagccgcatgtaataatgacacgaagaagaacagcgctcttaacggaagaataacat 4810
Dh 998 gtcagccgcatgtaataatgacacgaagaagaacagcgctcttaacggaagaataacat 1057
Qy 4811 gcgcgtcttctgacactctgacacacaaatgctgtaagaanaacagtcgtccagaata 4860
Dh 1058 gcgcgtcttctgacactctgacacacaaatgctgtaagaanaacagtcgtccagaata 1107

RESULT 11
AAK25593
ID AAK25593 standard; cDNA: 1140 BP.
AC AAK25593;
DT 02-AUG-1999 (first entry)
XX
DE Merozoite surface protein MSP-1-42 modified cDNA.
XX
KM MSP-1: merozoite surface protein; malaria; vaccine;
KM protein engineering; protein expression; codon usage;
KM transgenic animal; mutant; ss.
XX
OS Plasmodium falciparum.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..1131
FT sig_peptide 1..45
FT mat_peptide 46..1128
FT /*tag- b
FT /*tag- c
XX
PN MO9920766-A2.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98MO-US22225.
XX
PR 15-MAY-1998; 98US-0085649.
PR 20-OCT-1997; 97US-0062592.
XX
PA (GENZ) GENZYME TRANSGENICS CORP.
XX
PI Chen LH, Meade H;
XX
DR WPI: 1999-302742/25.
DR P-PSDB: AAY05834.
XX
XX
XX New modified recombinant nucleic acid sequences useful for producing

PT material DNA vaccine
XX
XX Claim 12; Fig 11; 43pp; English.
PS This novel, modified nucleic acid encodes a 42 kDa C-terminal
XX part (see AAY05834) of malaria merozoite surface protein MSP-1
CC (MSP-1-42), an important target for the development of a vaccine
CC against Plasmodium falciparum. The nucleic acid sequence has been
CC modified compared to the native sequence of MSP-1-42 (see AAK25587)
CC such that the AT content has been reduced and 10 mRNA instability
CC motifs eliminated while maintaining the protein amino acid sequence.
CC In addition, a sequence encoding a 15-amino acid beta-casein signal
CC peptide has been added to the 5' end of the sequence, and N262Q
CC and N181Q mutations have been introduced to eliminate
CC N-glycosylation sites. These alterations allow MSP-1-42 to be
CC expressed in the mammary gland (i.e. milk) of transgenic mice. The
CC invention provides modified recombinant nucleic acid sequences and
CC methods for increasing the mRNA levels and protein expression of
CC proteins that are difficult to express in cell culture systems,
CC mammalian cell culture systems or in transgenic animals. The
CC preferred difficult protein candidates for expression are those
CC derived from lower organisms such as parasites, bacteria and
CC viruses that have DNA coding sequences of high AT content or which
CC have mRNA instability motifs or rare codons relative to the
CC recombinant expression system to be used. The invention allows
CC expression of MSP-1 in the milk of transgenic animals, and also
CC provides a DNA vaccine comprising a vector containing the altered
CC MSP-1-42 sequence.
XX
SQ Sequence 1140 BP; 353 A; 282 C; 290 G; 215 T; 0 other;

Query Match 15.0%; Score 742; DB 20; Length 1140;
Best Local Similarity 80.8%; Pred. No. 1.2e-170;
Matches 865; Conservative 0; Mismatches 205; Indels 0; Gaps 0;
Qy 3791 ccggtgaagcgtcactccctcgtgattgataacattcttccaanaacgagacgaat 3850
Dh 38 ccaatgcagccgtcactccctcgtgattgataacattcttccaanaacgagacgaat 97
Qy 3851 aagaagtcgtctatctgaacactctgacgagctctataggtctctcaagaanaacgtcg 3910
Dh 98 acgagtgctgactgagacgcccctgacgagctctataggtctctcaagaanaacgtcg 157
Qy 3911 agaataacgtgatacctcaatgtaacgtggaagacattctcaaacgcccgttaata 3970
Dh 158 agaataacgtgatacctcaatgtaacgtggaagacattctcaaacgcccgttaata 217
Qy 3971 agagagaataattcaagaacgtcttgagagagacttatccctataaagaactgacct 4030
Dh 218 agagagaataattcaagaacgtcttgagagagacttatccctataaagaactgacct 277
Qy 4031 cctctaactacgttgcacaggaaccatcaagttctcctaataaagaagaggaataat 4090
Dh 278 gcagcaactacgttgcacaggaaccatcaagttctcctaataaagaagaggaataat 337
Qy 4091 ttctgtcagttacaactatacaagaactccatcgacacacgataataatttcgtcaatg 4150
Dh 338 tcttgaagcagttacaactatacaagaactccatcgacacacgataataatttcgtcaatg 397
Qy 4151 atgtgtcgtggtattacaagaactctgagcgaanaaataaagttgacacttaccataa 4210
Dh 398 atgtgtcgtggtattacaagaactctgagcgaanaaataaagttgacacttaccataa 457
Qy 4211 aagaatatacaagataaagaagcggaatagtaaaaatctccctctctgataataca 4270
Dh 458 aagaatatacaagataaagaagcggaatagtaaaaatctccctctctgataataca 517
Qy 4271 tcgaaacctgtaagaagcagtgacacaaatcgacactcttcgttaatttcacactgagg 4330
Dh 518 tcgaaacctgtaagaagcagtgacacaaatcgacactcttcgttaatttcacactgagg 577
Qy 4331 ccaaggtctctaactataactacagagaaggaatgtagtgagtgaaataaatacagaagcctga 4390

Db 961 gatgcgaatgtaccgaagaagattcaggtagcgaacggaagaatcacatgtaagtc 1020
OY 4819 actaagcccgactctatccactctctcgaaggatttttcg 4860
Db 1021 actaacctgattcttaccactcttcgattgatttttcg 1062

RESULT 14
AAK25587
ID AAK25587 standard; CDNA: 1088 BP.
XX
AC AAK25587;
XX
DT 02-AUG-1999 (first entry)
XX
DE Merozoite surface protein MSP-1-42 cDNA.
XX
KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal; ss.
XX
OS Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT CDS 1..1086
FT /*tag= a
XX
XX MO9920766-A2.
XX
XX 29-APR-1999.
XX
XX 20-OCT-1998; 98MO-US22225.
XX
XX 15-MAY-1998; 98US-0085649.
XX 20-OCT-1997; 97US-0062592.
XX
XX (GEN2) GENZYME TRANSGENICS CORP.
XX
PI Chen LH, Meade H;
XX
XX WPI: 1999-302742/25.
XX
XX P-PSDB: AAY05833.
XX
XX New modified recombinant nucleic acid sequences useful for producing
XX malarial DNA vaccine
XX
XX
XX Disclosure: Fig 2; 43pp; English.
XX
XX This nucleic acid encodes a 42 kDa C-terminal portion (see AAY05833)
XX of malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
XX important target for the development of a vaccine against
XX Plasmodium falciparum. The 3' end of the sequence has been
XX modified to include a 6xHis tag. The nucleic acid sequence has
XX been modified (see AAK25586) according to a method of the invention
XX in order to improve expression in mammalian host cells and in
XX transgenic animals. In the modified coding sequence, 306 nucleotide
XX positions have been replaced to lower the AT content (from 76 to
XX 49.7%) and to eliminate 10 mRNA instability motifs. The encoded
XX amino acid sequence is unaltered. In another modified sequence
XX (see AAK25593), a signal peptide sequence has been added and two
XX N-glycosylation sites eliminated. The invention provides modified
XX recombinant nucleic acid sequences and methods for increasing the
XX mRNA levels and protein expression of proteins that are difficult
XX to express in cell culture systems, mammalian cell culture systems
XX or in transgenic animals. The preferred difficult protein
XX candidates for expression are those derived from lower organisms
XX such as parasites, bacteria and viruses that have DNA coding
XX sequences of high AT content or which have mRNA instability motifs
XX or rare codons relative to the recombinant expression system to be
XX used. The invention allows expression of MSP-1 protein in the milk
XX of transgenic animals, and also provides a DNA vaccine comprising a
XX vector containing the altered MSP-1-42 sequence.

XX
SQ Sequence 1088 BP; 454 A; 139 C; 150 G; 345 T; 0 other;
Query Match 12.7%; Score 625.2; DB 20; Length 1088;
Best Local Similarity 74.3%; Pred. No. 3e-142;
Matches 789; Conservative 0; Mismatches 273; Indels 0; Gaps 0;
OY 3799 gctgcactctcttcggtatgatatacattcgtccaaatcgcgaagaatcgaagt 3858
Db 1 gaagtaactcctccgtatgtatatacactcttcaaatgtgaatgaatcagaagtc 60
OY 3859 ctctatcgaaacctctgcagcgctcatagtcgtcccaagaacagctggaatcac 3918
Db 61 ttattttaaaccttgcagcggttttatgaaagttcaaaaaaacattgaaataac 120
OY 3919 gtatgaccttcaatgtcaacgtggaagacattctgaacagcgcttataaagagaa 3978
Db 121 gtatgacatttaagttaagttaagatattttaattcagattcaataaagtgaa 180
OY 3979 aattcagaagcgtcttgagagcgacttgattcccttaagaagcctactccttac 4038
Db 181 aattcaaaaatgltttagatcagattcaattcaataaagatttaacatcaagtat 240
OY 4039 taagtgtcaaggaaccatacaagtlcctcaataaagaagaaggaatctcgtct 4098
Db 241 tatgtgtcgaagaagtcacataaatttcttaataaagaagaagataattcctaagc 300
OY 4099 agttacaactatatacgaactccatcgacacagatatcaatttcgtaatgattgctg 4158
Db 301 agttataattatgaagatccaatagatacgaataaatttgcgaatgatttctt 360
OY 4159 ggtattcaagatctcigagcgaataatcagaagtccttgacttgatcattaaagat 4218
Db 361 ggaattataaataattatccgaaataatataaataagattagattcaataaataat 420
OY 4219 atcaacgataagcgaagcgagaaatgaataatatactgccttcctgaataacatcgaaac 4278
Db 421 atcaacgacaacaaagcggaatgagaatactccctctttaaacaacattgagacc 480
OY 4279 ctgtacaagacagtgaacgacaataatcgactcttcgttaattcacctggaagccaagtc 4338
Db 481 ttataataaacagttaatgataaattgattatttgaattcatcattgaagcaaaagt 540
OY 4339 ctcaactacttaacgagaagaagcaatgtggaatgaatgaagaagcgtgaactc 4398
Db 541 ctcaattatacatagagaatccaacgtgagaatgaataaataaagaacttaattactta 600
OY 4399 aaacaatccaagacaagctgtgcagatttcaagaanaataaacaattcgtcgaattgca 4458
Db 601 aaacaatccaagacaagctgtgcagatttcaagaanaataaacaattcgtcgaattgct 660
OY 4459 gactcgtctacgattataaccacaacatctcctgaccgaagttctgtccaatgcatg 4518
Db 661 gattatcaacagattataaccataataactatgtgacaagaagttccttagtaacagtatg 720
OY 4519 gtgttcgaaacctgcgaacaaacagtgctggaattcgtcgcgaagaagcctggaagcc 4578
Db 721 gttttgaaaatctgtgcgaacaaacgctttatcttaattcattgaggaacttgaaagt 780
OY 4579 atgtcgaacatctcccaacacaatgctggaagaagaagtcgcccaaatagcgcgtg 4638
Db 781 atgttaaacattccacaacacaatgctgtaaaaaaacaatgtccacaacattcggatgt 840
OY 4639 ttcaagcatctgacgagcggaagagtgcagagtctcctcgtgaatacaacaagaagga 4698
Db 841 ttcaagcatttagatgaagaagaagatgtaattgttttaattcaacaacaagaagt 900
OY 4699 gataagtcggtggaagaaacccaacctcctgcgaatgaagaagaaggggtgtgagcc 4758
Db 901 gataaattgttgtaaaatccaatcctactctgtgaacgaataatggtgtgtgattgca 960
OY 4759 gatgtaatgtcacccgaggaagacagcgctcctaacgaaagaatacatatgagagt 4818

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2001, 13:12:16 ; Search time 136.83 Seconds
(without alignments)
6834.736 Million cell updates/sec

Title: US-09-269-874-2

Perfect score: 4940
Sequence: 1 cgcacgcgtatgaatcatc.....ttcatctaatagatcgtatg 4940

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backlilist.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1265.2	25.6	5181	1 US-08-257-073-10	Sequence 10, Appl
2	74.2	1.5	7218	1 US-08-232-463-14	Sequence 14, Appl
3	60.6	1.2	5163	3 US-08-700-651-1	Sequence 1, Appl
4	60.6	1.2	5163	3 US-08-928-361B-4	Sequence 4, Appl
5	60.6	1.2	5318	3 US-08-700-651-2	Sequence 2, Appl
6	60.6	1.2	5318	3 US-08-928-361B-3	Sequence 3, Appl
7	58.8	1.2	5511	3 US-08-928-361B-2	Sequence 2, Appl
8	58.8	1.2	7334	3 US-08-928-361B-1	Sequence 1, Appl
9	49.6	1.0	1430	2 US-08-276-452A-25	Sequence 25, Appl
10	49.6	1.0	1430	2 US-08-798-744-25	Sequence 24, Appl
11	48.2	1.0	1690	1 US-08-276-452A-24	Sequence 24, Appl
12	48.2	1.0	1690	1 US-08-798-744-24	Sequence 24, Appl
13	46.4	0.9	198	5 PCT-US95-10668-3	Sequence 3, Appl
14	46.4	0.9	198	5 PCT-US95-10668-4	Sequence 4, Appl
15	45.8	0.9	340	1 US-08-182-175A-104	Sequence 104, App
16	45.8	0.9	340	1 US-08-182-175A-104	Sequence 104, App
17	45.6	0.9	697	6 PCT-US92-06412-104	Sequence 104, App
18	45.6	0.9	1137	6 PCT-US92-06412-104	Sequence 104, App
19	45.4	0.9	198	5 PCT-US95-10668-1	Sequence 1, Appl
20	45.4	0.9	198	5 PCT-US95-10668-2	Sequence 2, Appl
21	44.8	0.9	2223	1 US-08-257-073-4	Sequence 10, Appl
22	44.4	0.9	4766	5 PCT-US93-07261-10	Sequence 2, Appl
23	44.2	0.9	2277	1 US-08-676-967-2	Sequence 2, Appl
24	44.2	0.9	2277	1 US-08-676-974-2	Sequence 2, Appl
25	44.2	0.9	2277	1 US-09-098-487-2	Sequence 2, Appl
26	43	0.9	1241	1 US-08-471-033-39	Sequence 39, Appl
27	43	0.9	1241	1 US-08-471-033-42	Sequence 42, Appl

28	43	0.9	1241	2 US-08-471-044-39	Sequence 39, Appl
29	43	0.9	1241	2 US-08-471-044-42	Sequence 42, Appl
30	43	0.9	1241	2 US-08-463-483A-39	Sequence 39, Appl
31	43	0.9	1241	2 US-08-463-483A-42	Sequence 42, Appl
32	43	0.9	1241	2 US-08-471-046A-39	Sequence 39, Appl
33	43	0.9	1241	2 US-08-471-046A-42	Sequence 42, Appl
34	43	0.9	1241	2 US-08-470-566B-39	Sequence 39, Appl
35	43	0.9	1241	2 US-08-470-566B-42	Sequence 42, Appl
36	43	0.9	1241	2 US-08-469-334-39	Sequence 39, Appl
37	43	0.9	1241	2 US-08-469-334-42	Sequence 42, Appl
38	43	0.9	1241	3 US-09-300-529-39	Sequence 39, Appl
39	43	0.9	1241	3 US-09-300-529-42	Sequence 42, Appl
40	43	0.9	1358	1 US-08-471-033-45	Sequence 45, Appl
41	43	0.9	1358	1 US-08-471-044-45	Sequence 45, Appl
42	43	0.9	1358	2 US-08-463-483A-45	Sequence 45, Appl
43	43	0.9	1358	2 US-08-471-046A-45	Sequence 45, Appl
44	43	0.9	1358	2 US-08-470-566B-45	Sequence 45, Appl
45	43	0.9	1358	2 US-08-469-334-45	Sequence 45, Appl

ALIGNMENTS

RESULT 1
US-08-257-073-10
Sequence 10, Application US/08257073
Patent No. 5766597
GENERAL INFORMATION:
APPLICANT: Paoletti, Enzo
APPLICANT: de Taisne, Charles
APPLICANT: Tine, John A
TITLE OF INVENTION: MALARIA RECOMBINANT POXYVIRUS VACCINE
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Curtiss, Morris & Safford, P. C.
STREET: 530 Fifth Avenue, 25th Floor
CITY: New York
STATE: New York
COUNTRY: UNITED STATES OF AMERICA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/257,073
FILING DATE: 09-JUN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/075,783
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/852,305
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/672,183
FILING DATE: 20-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2570
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066 CURTMS
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 5181 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-257-073-10

OY	2011	aagaaagagatagcaaacatcgaaagtgttcaatgcaccaagtcgagagcctgtacaagaa	2070
Db	2098	AAAAAGAAAGTAGTAAATTAAAGAAATTTATACCAAAAGTTTAAAGACATGTTAA--GA	21555
OY	2071	gagaaagaaacatctaaacttgaaagacgtcaagataactccgagccttccacgaaaga	2130
Db	2156	AAGAAACAGCTGCTTATTCAGATATTACACACTTTAATTCACGCAAGCGAAACACTG	22155
OY	2131	gagaataacgagacggtctacaccgaagcccgagcaacagcgcggtgtcagcttcgaagc	2190
Db	2216	AAGAT--GGGGGTCTCCACACACACACTTTATTOCCATCAGGAAACAGAAAGTAAACAA	2274
OY	2191	gatacggtgcaagctcaagcaagaagcagaagcagcagcctccagtgccagtgcc	2250
Db	2275	GAAACAGAGAAACAGAAAGAAACAGTAGAGCACACAAACACCGTAAACATTAATACCA	2334
OY	2251	gtccagagagctaaagctcaagtgctcacaaccaaccagcctcgtaataaagaacgag	2310
Db	2335	CCAAAGAGATAAAGTTGTTGAAAATTCATTAAGCAATTAAGATTAATGACATTCACAA	2394
OY	2311	aatgtaacgaactggaactaccttgagaagctctatgagttccgtatataactacatc	2370
Db	2395	GCCTTGACAAAAACAGTTTATCTAAAGAAATTAGATGATTTTAACTTAATCATATTA	2434
OY	2371	tgccaacaatatatcctcgctctccacaacactatgaacgagaagattcttaacagtac	2430
Db	2455	TGCTATTAATATTTATTTAGTATCAAACTGATAGTGACCAAAATTTATAGAGTATAT	2514
OY	2431	aagataaccaaagagaagagagatlaaacgctcctctgtgataccactgagcctgtgtc	2480
Db	2515	AATCTTACTCCAGAGAGAAATGAATTTAAATCAATGATCCATTTGATTTATTAATTT	2574
OY	2491	aatacagaacaacatctccgctatgtaattctatgttcgaatagcctacaactcttc	2550
Db	2575	AATTTTCAAAATATACATACCTGCTATGTATTCATTTATGATAGTATGACATATTTTA	2634
OY	2531	tctcaactgttcatatgtagatataatggaagagagatggtctgcaacctgtataactcaa	2610
Db	2635	CAACATCTCTTTTTGAAATTAATCAAAAGAAAGATTTATTTACATTAACCTAATAA	2694
OY	2611	gacaagcacaagatlaagaacctctcggaagagcgaagctgaaggtctc-----	2657
Db	2695	GAGGAATTCATCATCAAAAATTTATTAGAGACCAAAACAATTAATGGAACATCATCT	2754
OY	2658	-----caacctgttaaacactctctc	2681
Db	2755	ACATTCAGTCTCGAATTAACAAACGTAATTACTGCTCATTCGCAATCCGAAATCCAGTAATCC	2814
OY	2682	cagctcacaatgaacacctgtctctcaaccctcaagcaa-----	2720
Db	2815	CAAAACCAAAATCAATGATGATCCTCTCAACATTAACCAAAATGGTGTAGCTGATCATCT	2874
OY	2721	-----ggccgaagtgagcgtctlaacgcgcgaacctctcaactgaccaacttaataactca	2775
Db	2875	GGTCTCTGTAGTTGAAGAAATGATGATCCCTTAACAGTATGTTCTTAATTAAGTAACAT	2934
OY	2776	ctgaagactgtttagaaacatccgtctctccgcgaagaataaagaacat---ctaccaagaa	2832
Db	2935	TTGAAAGGTATTTGTATGTCCTCTTAATCTTTGGAAATTAACAACTTAAGTACATTAATCATTA	2994
OY	2833	cttatgagacaagaatcgtccgaagaacctctcagaaagatactgaaagacagcgacaca	2892
Db	2995	ACCATTTCTCAACACAGATGAGTAAATAATTTTATGAGAAATATTTTAAATAATATGATATCC	3054
OY	2893	ctctataagagagctctcctaactctcggaatactlaaagccgatatataactctct	2952
Db	3055	TATTTTATATGATATCAACAAATTCGTAAATATCTAAATTCAAAAGTAAATTAACAGTTTG	3114
OY	2953	aacgatatactlaaagctlaaagcctggaagagacatcaataaagcttgaagaaagacgtg	3012
Db	3115	ACC---GAAACACAAAAAATTCGATTAATTAATGATGAATTTAAAAAATTAAAAAGATATCTTTA	3171
OY	3013	caactgagcttgacactgttacaacaagtacaactgaacttgaagacgtcttcgcagaag	3072

Db	3172	CAGTTATCATTTGGATTATATATATAAATAATATAAATTTAAATTTGATAGATTATTTAAATAG	3231
Qy	3073	aagaagacagtcggcgaagtataagatcaga tcaagaatltactctgcgaaggagac	3132
Db	3233	AAAAAGAACTTTGGCCAGACAAAATGCAAAATTTAAAAAAGCTTACTTTATTAAAAAGAACAA	3291
Qy	3133	cttgaagcaaacccaactcaactcgaacaaatccgaacacagtaactcgaagacttccagtg	3192
Db	3292	TTAGAAATCAAAATTTGAATTTGCATTTAATACCCCATATATGTATTACAAACTTTTCGT	3351
Qy	3193	ttcttcaacaaagaagaagagccgagatcgcgcgagacagagaacactctgagaaacc	3252
Db	3352	TTCTTTAAACAAAAAAGAAAGCGTGAATAGCGAAGACGAAACACATTTAGAAAAACAA	3411
Qy	3253	aagatctcttcaaacactcaaaagcctcgaagtataataagggaggtcttctct	3312
Db	3412	AAATATATTTATTTGAACATTTATAAGGACTTTGTTAAATTTATTAATAGTGAAATCATCTCCA	3471
Qy	3313	ctgaagactctctccgagggagagatccagaccgaagataactaaccgaagccctgaagac	3372
Db	3472	TTAAAACTTTTAAGTGAAGATATCATTTCAACAGAAAGATATATTATGCCATTTAGAAAA	3531
Qy	3373	ttcaagtcctgcttaagctcgaaggcaagctgaagacaactcgaacctgagagaagag	3432
Db	3533	TTTACAGATTAAAGTAAATAATAGATGAAACCTCAATGATTAATTTACATTTAGAAAGAA	3591
Qy	3433	aagctcagctaccctcttagcgagcagtcacactgaatcgcgcgagctcaaggagatcat	3492
Db	3592	AAATATTTCTTTCTTATCATCAAGGGATTACATCATTTATTAATTCGAAATTTAAAGAAATTAFA	3651
Qy	3493	aagaacaagaactcacccggcgaatagcccaagcgaaataatactacagcgtgaatacga	3552
Db	3652	AAAAATAAATTTATACAGGTAATTTCTCCACAGTGAATAATATAGAAAGTTATACCAAGCT	3711
Qy	3553	ctggaacttactaagaagatctctgctcgtgaagaaacagatgctgcacgtggtgtc----	3608
Db	3712	TTAAATTTCTTCGAAAAATTTTCTCCGAGAAAGCAAAAGTTACAAAGTTGTAATCTCAACT	3771
Qy	3609	-----tgaactgtgctccgacacacgaggagagcttca	3661
Db	3772	CACACAGATGTAACTCCATCTCCATTAATCTGTAAAGGTAAAGTGTAAGTTACAGATCCACA	3831
Qy	3642	actcaagaag-----cctgcatctactcaatgctcgagccgagttcaatacaatcac	3693
Db	3832	AAAGAGGAACACAAATATCCAACTTCAGCGCTTTTATTACGAATTTACACAACTATATA	3891
Qy	3694	acatctcagaaactcgagatlgagatcgatgacgttcatcatgtgctactcttcggcgag	3753
Db	3892	CAATTACAAAAATTTATGACGAAGAAAGATGATCTCTTGTGTTATCCCATTTTGGAGAA	3951
Qy	3754	agcgagaggagactcagatgaccccgcccgaggtggttaccgggtgagagctgtcaactcttc	3813
Db	3952	TCCGAGATTAATGCGAATATTTTAGATCAAGTAAGTAATCGGGAAGC---AATATCTGTC	4008
Qy	3814	gtgatgtataactctgctccaatccgagaaacgaatacgaagtgtctatactgaacct	3873
Db	4009	ACAAATGATATATCTCTCTCAGGATTTGAAAAATGAATATGATGTTATATTTAAACCT	4068
Qy	3874	ctggaagcgtctatagtgctctcaagaaacagctlgagaaataacgtgataccttcaat	3933
Db	4069	TTAGCTGAGATATATATACACTTTAAAAAAAACAATTTGAAAAAACATTTTACATTATAT	4128
Qy	3934	gtcaacgtgaaagacatctcgaaacgcgcttataaagaagaaatltcaagaagctc	3993
Db	4129	TTAAATTTTGAACGATATCTTAAATTTACGCTTTTAAACAAAGAAATATTTCTTAGATGTA	4188
Qy	3994	ttgagagcgagactgtatctcttaagaagccbaactctcctaactacgctgttcaaggac	4053
Db	4189	TTAGATCTGATTTAATGCAATTTAAACATATATCTCAAAATGAATACATTTATTTGAAGAT	4248
Qy	4054	ccatacaagttcctcaataaagaagaggatlaaatltctgtctcagttcaactatactc	4113

D	4249	TCATTTAAATATTGGAATTTCAGACACAAAAAACACACTTTTAAAAAGTTACAAATATATA	4308
OY	4114	aaggaactccatcgacacccgatatacaatttcgcataatgctggtggtattacaagatc	4173
D	4309	AAACAATACGATGACAAATGATGATTAAATTTCGACAGGAAGGATATAGTTATGAAAG	4368
OY	4174	ctgagcgaaaaatacaaatgcttgaccccttgctctattaaagtatatataaagataag---	4230
D	4369	GTTTATACGAAATATATAGATGATTTAGATCAATTTAAAAAGTTATCAAGAAAGAAAG	4428
OY	4231	-----caa	4233
D	4429	GAGAGTTCCCATCATCATCCACCACCAACAACCTCCGTACCGACGAAAAACGACGACACAA	4488
OY	4234	ggcgcagatgaataaataatactgcctccttcgtaataacatcgaaacccctgacagaacatg	4293
D	4489	AAGAAGGAATGATGATGCTTCATTTTAAACAACATGTGACAGCTTATACATATACCTTA	4548
OY	4294	aagacacaaatcgaccccttcgttaattcaactcgtgagccgaaggtccctaactactac	4353
D	4549	GTTTATTAATTAATGACGATTCCTTAATTAATTAATTAATTAATTAATTAATTAATTAAT	4608
OY	4354	gagagaagcaatgctggaagttaaaatacaaggcgacgaactactactaaacaactcaaac	4413
D	4609	GAAATAAGATGATGACACATGTTTAAATAATACCTTAACCTTAGTGATTTTAAAGCAATGATGAC	4668
OY	4414	aagctgcgcagatcctcaagaaaaataaacaatttcgttcgaattgcagacctctcaacgat	4473
D	4669	AAATATATCTTTTAAAAACCATTAACGATTCGAAAGCATTTAAAAAATTTGATTAATGAT	4728
OY	4474	tataacacaacaactcctcctgaccaaagttctgccaatgacatgcatggtctgnaaaacctc	4533
D	4729	GATTCGAAAAAAGATATGCTTGGCAAAATTACTTAGTACGAGATGAG---TTCAAAAATTTT	4785
OY	4534	gcccaaacagctgctggaacatcgcgcgcagcgcaacctgcgaagcatgctgaatcctcc	4593
D	4786	CTTATATCAATTAATATCAAAATTAATTTGAAGGAAATTTCCAAAGATATGTTAAACCTTTTCA	4845
OY	4594	cagcacaactgcctggaagaacagctgccccagaatagcgcctgtctcaggaatctggac	4653
D	4846	CACACCAATGCGCTAAAAAACAATGCTCCAGAAATTTCTCGATGTTTCAGACATTTAGAT	4905
OY	4654	gagcgcgaagatgcaagctgctcctgaaactaaacaagaagaagagataagctgcgtgag	4713
D	4906	GAAAGAGAAAGATGTAATGTTTATTAATTTCAAAACAAGAGAGTGATTAATGCTTTGAA	4965
OY	4714	aaccacaacctcctcgcgaatgaaacaactgacggtgctggaagccgatgctcaatgaccc	4773
D	4966	AATCCAAATCTCTACTGTATAACGAAATTAATGAGTGATGATGACAGATGCCAAAGATATGC	5025
OY	4774	gaggaagacacgcgcctctcaacggaagaagaatacaactgagatgctactcaagcccaatcc	4833
D	5026	GAAACAATTCAGGTGACGATGCAAGGAAAGAAATTCACATGTGATGTATCAAACTGATTC	5085
OY	4834	tatccactctcgacgggaatttctgctccagctcaattctctggtgacatcctcctcgt	4893
D	5086	TATCCACTTTTCATGATGATTTCTTCGACGTCCTCTAATCTTCTGAGATATCATTTCTTA	5145
OY	4894	ctgatctcatgctcgatccctgtaacagctcatcctaa	4929
D	5146	TTAATACCATGTTAATATATATATACCTTTTCAATTTTA	5181

RESULT 2
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52

[illegible]

[illegible]

RESULT 5
 US-08-700-651-2
 Sequence 2, Application US/08700651B
 Patent No. 6015882
 GENERAL INFORMATION:
 APPLICANT: PETERSEN, CAROLYN
 APPLICANT: LEECH, JAMES
 APPLICANT: NELSON, RICHARD, C.
 APPLICANT: CUT, JIRI
 TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
 TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
 FILE REFERENCE: 480.19-4(HV)
 CURRENT APPLICATION NUMBER: US/08/700,651B
 CURRENT FILING DATE: 1997-08-14
 EARLIER APPLICATION NUMBER: 08/415,751
 EARLIER FILING DATE: 1995-04-03
 NUMBER OF SEQ ID NOS: 15
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 2
 LENGTH: 5318
 TYPE: DNA
 ORGANISM: Cryptosporidium parvum
 US-08-700-651-2

Query Match	Similarity	1.2%	Score	60.6	DB 3	Length	5318:
Best Local	Similarity	44.2%	Pred	No	6.7e-06		
Matches	249:	Conservative	0:	Mismatches	314:	Indels	0:
						Gaps	0:
QY	3930	caatgcaacgtgaaggacatcttgaacgcgcgtttaaagaagaaattccaaga	111	111	111	111	111
Db	615	caacacacaacaacagcaacaacaacactactaactactacactactactaga	111	111	111	111	111
QY	3990	cgcttgagagagcgcttgattccctatataagactcgactcctctaactgttca	111	111	111	111	111
Db	675	caacaacaacaacaacaacaacaacaacaacaacaacaacaacaaca	111	111	111	111	111
QY	4050	ggaccatacaagtttcctcaatcaagaagagaggataattctgtcagttaacata	111	111	111	111	111
Db	735	cgactactactactactactactactactactactactactactactaacaacaacata	111	111	111	111	111
QY	4110	tatcaaggatccatcgacacacagatatacaattcgtaatgctggtggtaltaca	111	111	111	111	111
Db	795	caaccaacaactcaacaacaacaactcaaccaacaacaacaacaacaacaactaca	111	111	111	111	111
QY	4170	gattctgagcgaaaatacaagcttgactctgactctataaaggtatataacagtaa	111	111	111	111	111
Db	855	ccacaactacaaccacaacaactacaaccacaacaacaaccacaaccacaactacca	111	111	111	111	111
QY	4230	gcaagcgagaaatgaaaataatctgcctctctgataatacatcgaaacctgtacaagc	111	111	111	111	111

Db	915	agaaacccaacacacacactactacaacaacaacaacacacacacactactactacaaca	974
Oy	4230	agtgaaagcaaaaatcgaccttctgtaattcaactgagggcnaagttccttaactaac	4339
Db	975	ccacaacacacacacacacacacacactacaactcccaagaacacacactactacta	1034
Oy	4350	ttacgagaagagcaatgtggaagtttaaatcagagagtcgaaactcccaaacatcca	4409
Db	1035	ctaccacaacacacacacactactactactaccacaacacacacactactactacta	1094
Oy	4410	agacaagctgycagattcacaagaataaacaaattcgtcggaaatlbgagacctgtctac	4469
Db	1095	caacacacacacacacacacacacacacacacacactaccggaacacacacacacaa	1154
Oy	4470	cgattataacacacacatctcc	4492
Db	1155	caactactactactactacaacc	1177

```

1      RESULT      6
2      US-08-928-361b-3
3      ; Sequence 3, Application US/08928361B
4      ; Patent No. 6071518
5      ;
6      ; GENERAL INFORMATION:
7      ; APPLICANT: Petersen, Carolyn
8      ; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
9      ; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
10     ; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
11     ; TITLE OF INVENTION: SPECIES INFECTIONS
12     ;
13     ; NUMBER OF SEQUENCES: 30
14     ;
15     ; CORRESPONDENCE ADDRESS:
16     ; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
17     ; STREET: 385 Sherman Avenue, Suite 6
18     ; CITY: Palo Alto
19     ; STATE: CA
20     ;
21     ; COUNTRY: USA
22     ; ZIP: 94306-1840
23     ;
24     ; COMPUTER READABLE FORM:
25     ;
26     ; MEDIUM TYPE: Floppy disk
27     ; COMPUTER: IBM PC compatible
28     ; OPERATING SYSTEM: PC-DOS/MS-DOS
29     ; SOFTWARE: PatentIn Release #1.0, Version #1.30
30     ;
31     ; CURRENT APPLICATION DATA:
32     ; APPLICATION NUMBER: US/08/928,361B
33     ; FILING DATE: 12-SEP-1997
34     ;
35     ; CLASSIFICATION:
36     ;
37     ; PRIOR APPLICATION DATA:
38     ; APPLICATION NUMBER: US 60/026,062
39     ; FILING DATE: 13-SEP-1996
40     ;
41     ; ATTORNEY/AGENT INFORMATION:
42     ;
43     ; NAME: VERNY, Hana
44     ; REGISTRATION NUMBER: 30,518
45     ; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
46     ;
47     ; TELECOMMUNICATION INFORMATION:
48     ; TELEPHONE: 650-324-1677
49     ; TELEFAX: 650-324-1678
50     ;
51     ; INFORMATION FOR SEQ ID NO: 3:
52     ; SEQUENCE CHARACTERISTICS:
53     ; LENGTH: 5318 base pairs
54     ; TYPE: nucleic acid
55     ; STRANDEDNESS: double
56     ; TOPOLOGY: linear
57     ;
58     ; MOLECULE TYPE: DNA (genomic)
59     ;
60     ; US-08-928-361b-3

```

	Query Match	1.2%	Score 60.6:	DB 3:	Length 5318:
	Best Local Similarity	44.2%:	Pred. No. 6.7e-06:		
	Matches 249:	Conservative	0:	Mismatches 314:	IndeIs 0:
Oy	3930	caatgtcaacgctgaagaagattctgaaccgcgcgtttaataagaagagaanaatlccaagaa	3989		
Dd	614	CAACAGACAACAAACGACGAACAAACAGCAACTCTACACTCTACACTCTACACTCTACCA	673		

```

OY 3990 cgtcttgagagcgactgtatccctataaagacgtgacctctcttaactcgtgtca 4049
Db 674 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 733
OY 4050 ggaccatacaagttcctcaataaagagaagaggaataattctgtctagtaacta 4109
Db 734 CGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 793
OY 4110 tatcaaggactccatcgaccgcgatatacttgcctaagtgtgcgggtataca 4169
Db 794 CAACCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 853
OY 4170 gatcctgagcgaataatacaagctgccttgcctctataaagaatatacaagata 4229
Db 854 CCACAAGTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 913
OY 4230 gcaagcgagagatcaaaaatactctgacctctctgaataacatcgaaacctgtacaagac 4289
Db 914 AGAAGACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 973
OY 4290 agtgaagcgaataatgacctctctgtaattcacttgagcgcaaggtcccaactatac 4349
Db 974 CCACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1033
OY 4350 ttacgagagagcgaatgtggaagttaaatcaagagctgtaactacccaacaacatca 4409
Db 1034 CTACCACAGACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1093
OY 4410 agacaagctgcagatttcaagaataatacaattcgtcggaattgcagacctgtctac 4469
Db 1094 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1153
OY 4470 cgattataaccaacaacatctcc 4492
Db 1154 CAACACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 1176

RESULT 7
US-08-928-361B-2
Sequence 2, Application US/08928361B
Patent No. 6071518
GENERAL INFORMATION:
APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESSES:
ADDRESSEE: PETERS, VERNY, JONES & BIKSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: VERNY, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480,76-1(HV)
TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5511 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-928-361B-2

```

```

Query Match 1.2%; Score 58.8; DB 3; Length 5511;
Best Local Similarity 45.2%; Pred. No. 2.1e-05;
Matches 216; Conservative 0; Mismatches 262; Indels 0; Gaps 0;

```

```

OY 3936 caacgtgagagacattctgaacgcgctttaaagagagaataattcaagaagctct 3995
Db 1094 CCAGCACAACACTACAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1153
OY 3996 ggagagcgactgtatccctataaagacgtgacctcttaactacgtgtcgaagacc 4055
Db 1154 CACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 1213
OY 4056 atacaagttccctcaataaagagaagaggaataattctgtctgtacactatatac 4115
Db 1214 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 1273
OY 4116 ggaactcagacacgcatatcaattcgtctaigtgtgtggtatcaagaatcct 4175
Db 1274 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1333
OY 4176 gaggaaataatacagttcgtccttgccttgccttataaagaatatacaagtaagc 4235
Db 1334 CAACCACAACCAACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 1393
OY 4236 cgagatgaataatactgcctctcctgaataacatcgaaacctgtacaagaagtga 4295
Db 1394 CAACAACAACACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 1453
OY 4296 cgacaataatcgacctctcgttaattcaccttgagagcgaaggtccctcaactacta 4355
Db 1454 CCACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACTA 1513
OY 4356 gaagagcagatgtgaagttaaatcaagagcgtgaactactcaaaacatccaagac 4413
Db 1514 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 1571

```

```

RESULT 8
US-08-928-361B-1
Sequence 1, Application US/08928361B
Patent No. 6071518
GENERAL INFORMATION:
APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESSES:
ADDRESSEE: PETERS, VERNY, JONES & BIKSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

```


Db 701 AACACAACAGGCTTCTCCGAGATTACACAGCAACACACAAATATTTTC 760
QY 3325 tccgaggagagcatccagaccgaggaatactacgcccagctccgagaacttaagtcctg 3384
Db 761 TCCGAGATTACACAAATATTAACATATATGTTTCTCCGAGATTACACAAATGCT 820
QY 3385 tctaagctcgaaggaagctgaaggaacacctggaacctggaagaagaagctcagctac 3444
Db 821 AATAACATATATGTTTCTCCGAGATTACACAAATATTAACAAATGCTTCTCCGAG 880
QY 3445 ctctcgtcggactgcatacctcgtatcgccgagctcagaagaatcattagaacaagaac 3504
Db 881 AACACAAACAATATATGTTTCTCCGAGATTACACAAATATTAACAAATGCT 940
QY 3505 tacaccggaatagcccaagcgaggaataatacagac 3540
Db 941 TTCTGTGAGATTACACAAACAAACACAAATGCC 976

RESULT 10
US-08-798-744-25
Sequence 25, Application US/08798744
Patent No. 5830747
GENERAL INFORMATION:
APPLICANT: Chen, Chao-Guang
APPLICANT: Mau, Shaio-Lim
APPLICANT: Du, He
APPLICANT: Gane, Allison M
APPLICANT: Bacic, Antony
APPLICANT: Clarke, Adrienne E
TITLE OF INVENTION: Plant Arabinoxylactan Protein (AGP) Genes
NUMBER OF SEQUENCES: 91
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: United States of America
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/798,744
FILING DATE: 13-FEB-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/276,452
FILING DATE: 18-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Caruthers, Jennie M.
REGISTRATION NUMBER: 34,464
REFERENCE/DOCKET NUMBER: 27-91A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303)499-8080
TELEFAX: (303)499-8089
TELEX: 49617824
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1430 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 2..1312
FEATURE:
NAME/KEY: misc_feature
LOCATION: 41..112
OTHER INFORMATION: /note= "Derived amino acid sequence

OTHER INFORMATION: corresponding to the peptide sequence by protein
OTHER INFORMATION: microsequencing"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 41..112
OTHER INFORMATION: /note= "Derived amino acids 14-24,
OTHER INFORMATION: 28, 30, 32-37 are identical to sequences obtained
OTHER INFORMATION: by protein microsequencing"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 25..31
OTHER INFORMATION: /note= "Amino acids 25, 27, 29, and
OTHER INFORMATION: 31 are hydroxylated proline residues; amino acid
OTHER INFORMATION: 26 can be T instead of A"
US-08-798-744-25

Query Match 1.0%; Score 49.6; DB 2; Length 1430;
Best Local Similarity 44.3%; Pred. No. 0.0025;
Matches 202; Conservative 0; Mismatches 254; Indels 0; Gaps 0;

QY 3085 ggcaagatataagatgcagatcaagaagctgctcgtcgaagagcgagcttgaagaaga 3144
Db 521 GGCTACTCGAGATTACACAAACAGCAACAAATGCTACTCGAGATTACACAAAC 580
QY 3145 ctcaactcagaacaatccgaacacgtaactgcagaactctcagtgcttcaacaag 3204
Db 581 AACACAACAATGCTCTCTCCGAGATTACACAAACAACAATGCTACTCCAGAT 640
QY 3205 aagaaggaagcgagatcgccgagagagacacactctggaacacacagatctctc 3264
Db 641 TACAAACAATGCTCTCTCTCCGAGATTACACAAACAACAATGCTACTCCAGAT 700
QY 3265 aaactacaaagagcctcgatataatataagtggaagcttctcctcgaagctc 3324
Db 701 AACACAACAAGGCTTCTCCGAGATTACACAAACAACAACAATATTTTC 760
QY 3325 tccgaggagagcatccagaccgaggaatactacgcccagctccgagaacttaagtcctg 3384
Db 761 TCCGAGATTACACAAATATTAACATATATGTTTCTCCGAGATTACACAAATGCT 820
QY 3385 tctaagctcgaaggaagctgaaggaacacctggaacctggaagaagaagctcagctac 3444
Db 821 AATAACATATATGTTTCTCCGAGATTACACAAATATTAACAAATGCTTCTCCGAG 880
QY 3445 ctctcgtcggactgcatacctcgtatcgccgagctcagaagaatcattagaacaagaac 3504
Db 881 AACACAAACAATATATGTTTCTCCGAGATTACACAAATATTAACAAATGCT 940
QY 3505 tacaccggaatagcccaagcgaggaataatacagac 3540
Db 941 TTCTGTGAGATTACACAAACAAACACAAATGCC 976

RESULT 11
US-08-276-452A-24
Sequence 24, Application US/08276452A
Patent No. 5646029
GENERAL INFORMATION:
APPLICANT: Chen, Chao-Guang
APPLICANT: Mau, Shaio-Lim
APPLICANT: Du, He
APPLICANT: Gane, Allison M
APPLICANT: Bacic, Antony
APPLICANT: Clarke, Adrienne E
TITLE OF INVENTION: Plant Arabinoxylactan Protein (AGP) Genes
NUMBER OF SEQUENCES: 91
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: United States of America

```

1 ZIP: 50303
2 COMPUTER READABLE FORM:
3 MEDIUM TYPE: Floppy disk
4 COMPUTER: IBM PC compatible
5 OPERATING SYSTEM: PC-DOS/MS-DOS
6 SOFTWARE: Patentin Release #1.0, Version #1.25
7 CURRENT APPLICATION DATA:
8 APPLICATION NUMBER: US/08/276,452A
9 FILING DATE: 18-JUL-1994
10 CLASSIFICATION: 435
11 ATTORNEY/AGENT INFORMATION:
12 NAME: Caruthers, Jennie M.
13 REGISTRATION NUMBER: 34,464
14 REFERENCE/DOCKET NUMBER: 27-91A
15 TELECOMMUNICATION INFORMATION:
16 TELEPHONE: (303)499-8080
17 TELEFAX: (303)499-8089
18 TELEX: 49617824
19 INFORMATION FOR SEQ ID NO: 24:
20 SEQUENCE CHARACTERISTICS:
21 LENGTH: 1690 base pairs
22 TYPE: nucleic acid
23 STRANDEDNESS: single
24 TOPOLOGY: linear
25 MOLECULE TYPE: CDNA
26 FEATURE:
27 NAME/KEY: CDS
28 LOCATION: 60..1442
29 FEATURE:
30 NAME/KEY: misc_feature
31 LOCATION: 1..38
32 OTHER INFORMATION: /note= "Nucleotide sequence
33 OTHER INFORMATION: obtained by PCR which does not overlap with the
34 OTHER INFORMATION: CDNA clone"
35 FEATURE:
36 NAME/KEY: misc_feature
37 LOCATION: 60..128
38 OTHER INFORMATION: /note= "Predicted transmembrane
39 OTHER INFORMATION: segment"
40 FEATURE:
41 NAME/KEY: misc_feature
42 LOCATION: 135..179
43 OTHER INFORMATION: /note= "Derived amino acid sequence
44 OTHER INFORMATION: corresponding to the peptide sequence by protein
45 OTHER INFORMATION: microsequencing"
46 FEATURE:
47 NAME/KEY: misc_feature
48 LOCATION: 135..179
49 OTHER INFORMATION: /note= "Amino acids 27 to 36, 38,
50 OTHER INFORMATION: and 40 are identical to that in the peptide
51 OTHER INFORMATION: obtained by direct microsequencing"
52 FEATURE:
53 NAME/KEY: misc_feature
54 LOCATION: 135..179
55 OTHER INFORMATION: /note= "Amino acid 26 may also be
56 OTHER INFORMATION: Ala; 37 and 39 can also be undetermined residues"
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
9
```

	Query Match	1.0%	Score 48.2	DB 1	Length 1690	
	Best Local Similarity	43.2%	Pred. No. 0.0065			
	Matches 230	Conservative	0	Mismatches 303	Indels 0	Gaps 0
OY	859	gaccgtccatcctataaacacagcttgaagaagcccaactcatcggtacgtagag	918			
Db	573	GATGAGTCTCAATAAACAATGGTTATGATTCCAAACAAACAGAACAAACAATGGTTAT	632			
OY	919	aagcgcatagacacctaaagaaagtgaanaalacaaagaactctgcacaagattaat	978			
Db	653	GATTCCACCAATATACAAACAATATACGATGTGGCTTCGCCAGATTTTCAAACAACAAT	692			
OY	979	gaatttaagaatcccgcgcacagccaactcttgtaaacaccctcaacagctgtgtgacaag	1038			

```

Db      693  GGCTACTCGGAGAAATGCTAATAACAAAAATTAACAATGCGCTACTCAGAGAAATTACAACAC 752
QY      1039  aacaagaagatagagagacccggaagaagatcaagaagatcgcgaataacattaaagtc 1098
Db      753  AATACACACATGGCTACGCCAAGAATTAACACAAATGGCTACTCTCAGATTACAAACAC 812
QY      1099  aacatagattctcctcttactgatacccttgagctcgagtaactcttgagagaagaat 1158
Db      813  AACAAATATTTTACTCTCGAGAAATTACACACACACACAAACAAATGTTTTCTCGAGAGAT 872
QY      1159  aagaatatagacatctccgcgaagtcgagacaaagaaatcaaccgaaccttaatgaatat 1218
Db      873  TCCAAACAACAATGGCTACTCTCAAAAAGATCAACAAATAAATGCTACTCCGAGAAATTACATG 932
QY      1219  cceaattgctgaagctacccctcgtcttataacgatatcaaaaagcctctaaagagtc 1278
Db      933  AACCAACACAAATGGCTTCTCCGAGGTTTCAACACACACACACACACACACACACAC 992
QY      1279  aatagctcggtactgattaaacccctcgattatagagaagaacccctaaagaatc 1338
Db      993  AACGTTTTCTGTGAATTAACAACAACAACATACAAATAAATGTTTTTCTCGAGAAATTAC 1052
QY      1339  tacacgacacaatgagagaagaattatcaaccgaatatcaagaagaagatcaa 1391
Db      1053  AACCAACACATATAACAACATGCTTCTACGAGAAATTACACACACACACACACACAC 1105

RESULT 12
US-08-798-744-24
Sequence 24, Application US/08798744
Patent No. 5830747
GENERAL INFORMATION:
APPLICANT: Chen, Chao-Guang
APPLICANT: Mau, Shiao-Lim
APPLICANT: Du, He
APPLICANT: Gane, Allison M
APPLICANT: Bacic, Antony
APPLICANT: Clarke, Adrienne E
TITLE OF INVENTION: Plant Arabidogalactan Protein (AGP) Genes
NUMBER OF SEQUENCES: 91
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: United States of America
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/798,744
FILING DATE: 13-FEB-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/276,452
FILING DATE: 18-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Caruthers, Jennie M.
REGISTRATION NUMBER: 34,464
REFERENCE/DOCKET NUMBER: 27-91A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303)499-8080
TELEFAX: (303)499-8089
TELEX: 49617824
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 1690 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```


APPLICANT: James Eberwine
TITLE OF INVENTION: A Method of Sequencing Proteins by
TITLE OF INVENTION: Epitope Ordering and Protein
TITLE OF INVENTION: Restriction Mapping
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jane Massey Licata, Esq.
STREET: 210 Lake Drive East, Suite 201
CITY: Cherry Hill
STATE: NJ
COUNTRY: USA
ZIP: 08002
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM 486
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10668
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/294,133
FILING DATE: August 22, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jane Massey Licata
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: PENN-0137
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 779-2400
TELEFAX: (609) 779-8488
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 198
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: No
PCT-US95-10668-4

Query Match 0.9%; Score 46.4; DB 5; Length 198;
Best Local Similarity 52.6%; Pred. No. 0.0054;
Matches 101; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
QY 1357 aagaagttatcaacgaatcaagaagatcaaatgtgagaagaagaatgtaggt 1416
DB 2 AACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 61
QY 1417 gacaagaagttacgaagcgcagcaaaagtctaacgatatacctaagaagatgaa 1476
DB 62 AACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 121
QY 1477 aagctgtgaacgagatctatgatctcaaatcaacaaataacatgacctgacaaattc 1536
DB 122 AACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 181
QY 1537 gagaagaatgatg 1548
DB 182 AAGCTGATGCTG 193

RESULT 15
US-08-182-175A-104
Sequence 104, Application US/08182175A
Patent No. 5559223
GENERAL INFORMATION:
APPLICANT: Saverio Carl Falco
APPLICANT: Sharon J. Keeler
APPLICANT: Janet A. Rice
TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing F
NUMBER OF SEQUENCES: 113
CORRESPONDENCE ADDRESS:

ADDRESSEE: E.I. du Pont de Nemours and Company
STREET: 1007 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: USA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word, 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/182,175A
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/743,006
FILING DATE: 9 August 1991
ATTORNEY/AGENT INFORMATION:
NAME: Linda Axamethy Floyd
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: BB-1031
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 992-4929
TELEFAX: (302) 892-7949
TELEX: 835420
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 340 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
STRAIN: E. coli
CELL TYPE: DHS alpha
IMMEDIATE SOURCE:
CLONE: segment 534 [seg 534]
FEATURE:
NAME/KEY: CDS
LOCATION: 3..326
OTHER INFORMATION: /function= "synthetic seed storage protein"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "ssp"
OTHER INFORMATION: /standard_name= "SSP-534"
US-08-182-175A-104

Query Match 0.9%; Score 45.8; DB 1; Length 340;
Best Local Similarity 49.0%; Pred. No. 0.011;
Matches 122; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
QY 2973 gaagctggaagaagacatcaataagctgaagaagaactgcaactgagctcgactgt 3032
DB 62 GAAGCTGAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAT 121
QY 3033 caacagtlacaacactgaacgagagagactctcgacaagaagaagaagtcggcaagta 3092
DB 122 GAAGAGCTGAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 181
QY 3093 taagatcagatcaagaagttgactctgctcaagaagcagcttgaagaacaactcaactc 3152
DB 182 GATGAAGAAGCTCGAAGAGAGAAAGAAAGCTCATGAGAGAGAAAGAAAGCTCGAAGA 241
QY 3153 actgacaatccgaagaacgactgcaagaacttctcagtgcttcaacaagaagaagga 3212
DB 242 AAGATGAAGCAATGAAGAGAGAAAGAAAGTGAAGTGAAGAGAAAGAGCTCGA 301
QY 3213 agccgagat 3221
DB 302 AGAGAGAT 310

Mon Oct 29 07:52:14 2001

us-09-269-874-2.rni

Page 13

Search completed: October 27, 2001, 13:13:25
Job time: 14001 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 27, 2001, 13:10:47 : Search time 8587.58 Seconds
(without alignments)
8897.813 Million cell updates/sec

Title: US-09-269-874-2
Perfect score: 4940

Sequence: 1 cgcacgcgatgataatcat.....ttcatctaatagatgatg 4940

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: gb_pl5:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_r01:*
95: gb_r02:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	4940	100.0	4940	9 A92451	A92451 Sequence 2
2	4938.4	100.0	4940	56 CVE131294	AT131294 Cloning v
3	2816.4	57.0	5917	96 PFANT195	X02919 Plasmodium
4	2816.4	57.0	7038	96 PFEP195A	X15063 Plasmodium
5	2814.4	57.0	4920	9 A92450	A92450 Sequence 1
6	2706.2	54.8	5760	9 A04562	A04562 P.falciparu
7	2706.2	54.8	5760	10 E00656	E00656 cdna encodl
8	2686.8	54.4	5282	96 PFP190	X03371 P.falciparu

9	1400.5	28.4	527.1	5	AF062349	Plasmodium
10	1331.2	26.9	523.1	5	AF062348	Plasmodium
11	1330.8	26.9	5085	5	AF218248	Plasmodium
12	1290.4	26.1	5392	96	PF0190G1	X05624
13	1290.4	26.1	5754	96	PF06MSGA	M19143
14	1272.2	25.8	5316	96	PF06MSA1B	Z35327
15	1267.2	25.7	5276	96	PF06GPI95	M73213
16	1267.2	25.7	5541	96	PF06P195	X08831
17	910.4	18.4	1636	96	PF06MSB06	P.falcipar
18	910.4	18.4	1636	96	PF06MSB08	D13357
19	908.8	18.4	1636	96	PF06MSB28	D13360
20	908.8	18.4	1636	96	PF06MSB37	D13362
21	908.8	18.4	1636	96	PF06MSB79	D13356
22	907.2	18.4	1636	96	PF06MSB834	D13361
23	907.2	18.4	1636	96	PF06MSB844	D13363
24	905.6	18.3	1636	96	PF06MSB8222	D13359
25	784.6	15.9	3395	96	PF06MSAS1	X61930
26	711.8	14.4	3518	96	PF06P190A	M57727
27	711.8	14.4	3518	96	PF06P190G	P.falcipar
28	691	14.0	1203	96	PF06MSD1	Y00087
29	635.8	13.3	1131	5	AF335919	Plasmodium
30	623.6	12.6	1065	96	PF06SURFPRO	AF335919
31	560.8	11.4	999	96	PF06MSA1AC	L20092
32	560.8	11.4	999	96	PF06MSA1AF	M77715
33	560.8	11.4	999	96	PF06MSA1AG	M77718
34	560.8	11.4	999	96	PF06MSA1AG	M77719
35	559.2	11.3	999	96	PF06MSA1AA	M77725
36	556.6	11.3	1133	5	AF06B1143	P.falcipar
37	547.8	11.1	1080	5	AF06B6876	AF06B1143
38	546.4	11.1	1956	96	PF06MSA1AD	Z35329
39	540.4	11.0	1956	96	PF06MSA1A	Z35326
40	540.4	11.0	1141	5	AF06B1141	P.falcipar
41	538.8	10.9	1142	5	AF06B1140	AF06B1140
42	536.8	10.9	999	96	PF06MSA1AH	M77720
43	535.4	10.8	1959	96	PF06MSA1C	Z35328
44	528.2	10.7	1116	5	AF06B1146	P.falcipar
45	527.8	10.7	1009	96	PF06B3A	AF06B1146

Qy	61	caagcgctgacccacgaatccctatccgggctctgtttaagaacttgaagat	120
Db	61	CAGGCGGAGACCCAGCAATCTATCAGAGACTGTTTAGAAACTGGAACTTTGGAAAT	120
Qy	121	ggcctccctaacggataacagccctctccaaagaagaagatgtgtctgaatgaaggagc	180
Db	121	GCCCTCTTACCCGGATATCAACCTGTTTCCAGAGAGAGACATGTGCTCTGAATAGAGACG	180
Qy	181	agtgcacgagccgtctacaacacagcacaccggctctctaaaggctctgtgactagcgtgtgc	240
Db	181	AGTGGCAAGGCGTTTACAAACGACACCCGGTTTAAAGGGTCTGTGCTTAGCGGTGGC	240
Qy	241	tccggctggtctctgtgaccttgggggtccctgcctccggcggcagcgttgacatcagt	300
Db	241	TCCGGTGGGTCTGTGGCCTCTGGGGGGTTCCTGGTCCCGCGGCGCTGTGCATCAGGT	300
Qy	301	ggctccatgtgcaagcagcgcgtctccgggaagaagtcggaagaaccaatccatctgaacaact	360
Db	301	GGCTCAGTGGCAACCGGGGCTTCCGGGAACATGTGAGAAGCAATCAATCAATCAACTT	360
Qy	361	agcattccagcagcagctctcagccgacccctaagacacagatgtgaacatcatctctc	420
Db	361	AGCCATTCCGACGCCCAAGTCTTACGCCGACCTCAAGCACCGAGTGGAAATATCTCTCTC	420
Qy	421	actatcaagagactgaagtaaccacagttgttcgacctactaatcaatlatatgtctgaactg	480
Db	421	ACTATCAAGAGGAGTGAAGTACCACCACTGTTTGACACTCTCATATCATATGTGTGACACTG	480
Qy	481	tgtataacatctatgctctccaatatctatctgaattgacggttaagaaagatcaatgaactc	540
Db	481	TGTATATACATCTATGGCTTCAATATCTATTTAGAGGTATTCACAAAGATGCATATGAACTC	540
Qy	541	ctgtacacagtgaattctctactctgacctgtcgaaggccaactgtgaatgaactgttggccc	600
Db	541	CTGTACAGTGTGAATTTCTACTTCTGACTTCTCTTAAGGGCCAAACTGATATACGTTTGGCCC	600

```

RESULT      1
A92451
LOCUS
DEFINITION  Sequence 2 from Patent WO9814583.
ACCESSION  A82451
VERSION     A92451.1 GI:6741181
KEYWORDS
SOURCE      unidentified.
ORGANISM    unidentified
REFERENCE   1 (bases 1 to 4940)
AUTHORS     Pan,W. and Bujard,H.
TITLE       METHOD FOR PRODUCING RECOMBINANTS INTENDED FOR USE IN A COMPLETE
            MALARIA ANTIGENE GP190/MSPI
JOURNAL     Patent: WO 9814583-A 2 09-APR-1998:
            PAN WEIQING (DE); BUJARD HERMANN (DE)
FEATURES
            Location/Qualifiers
            source          1..4940
                           /organism="unidentified"
                           /db_xref="taxon:32644"
BASE COUNT  1669 a 1157 c 1060 g 1054 t
ORIGIN
Query Match      100.0%: Score 4940; DB 9; Length 4940;
Best Local Similarity 100.0%: Pred. No. 0;
Matches 4940; Conservative 0; Mismatches 0; Indels 0; Gaps

```

D	b	601	AATGACATTGTCAAAATTCCATTCAATTGAAGATCAGAGCCAAACGATTGGACGATTG	660
O	y	661	aagaagttggtcttcggatcgcgaacgcccttcgcacaacatcaaggacaaagtgggaaag	720
D	b	661	AAGAAGTTGGTCTTCGGATTCTCCAAACCCTCTGCACACACTCMAAGGACAATGTGGGAAAG	720
O	y	721	atggaagattatatataaaaagaataagaaccatcggaacacttaagagcttgatcgaa	780
D	b	721	ATGGAAGATTATATAAAAAGAATTAACAAACCATCGAAGAACCTTAAACGATGATCGAA	780
O	y	781	gaatccaaaagaccatagacaanaaatlaagaatgcgaaccaagggaaagaaaaagaag	840
D	b	781	GAAATCCAAAAAGACCATATGACAAAAAATAGAAATGCACCAAGSAGSAGAAAAAGAAAG	840
O	y	841	tftgacaaagcccagtaagcacctgctccatclataaacaacgcttgaagaagcccataac	900
D	b	841	TFTGACAAAGGCCAGTAGACAGCTGTCATCTTATTAACAAACGCTTGAAAGGCCATTAC	900
O	y	901	ctcatcagcgtactcggagaagcgcatagacaacctcaagaagaatyaanaatatcaaaga	960
D	b	901	CTCATCACGCGTATCGAGGAAGCCCATGACACCCTCMAGMAGMTGMAAATATCAAAGAA	960
O	y	961	ctgtcgcacaagaattatgaanaatlaagaatccctccgcgcgcgaacactctggaagacccct	1020
D	b	961	CTGCTCGACAAGATTATGAAATTTAAGAAATCTCTCCGCAACCACTCTGGGAACACCCCT	1020
O	y	1021	aacacgctgcgtgcacaagaacaagaagatatagagagcaacgaaanaatgataccaagatc	1080
D	b	1021	AACACGCTGCTGCAACGAACCAAGMAGATAGAGACACGAGAAABAGATCCAAGAGATC	1080

Oy 1 cgcacgcgatgaaatcatcttcctccctcgtcatcttgltttatcatcaatcat 60
|||||
1 cgcacgcgatgaaatcatcttcctccctcgtcatcttgltttatcatcaatcat 60

Db 1081 GCCAAMACCTTAAGTCAACATAGATTCTCTTACTGATCCCCCTTGAGCGCTGGAGTAC 1140

OY	1141	tacttgtagagagaagaataaagaataaagacatctccgcgaagtcgagacaaaggaatca	1200
Db	1141	TACTTGGAGAGGAAAGAAATTAAGATTAAGACATCTCCGCCAAAGTGAGACAAAGGAATCA	1200
OY	1201	acggaacctaatgaataatcccaatggtgtagagatcccctctctataaagataaac	1260
Db	1201	ACCGAACCTTAATGAAATATCCCAATGGTGTGACATCCCTCTCTCTTATPAAGATATAC	1260
OY	1261	aagcgtctcaacgagctcaataagcttcggtgagcttgatataaacccttcgatatacgaa	1320
Db	1261	AAGCGTCTCAACGAGCGTCATPACTTCCGGTAGCTTATTAACCCCTTGATATTACGAA	1320
OY	1321	gaaccctctaaagaatactacacagcaatgtagaagaagttatcaacgaaatacaag	1380
Db	1321	GAACCCCTCTAAGAAATATCTCACAGCAATGTAGAGAAAGAAAGATTATACAGAAATCAAG	1380
OY	1381	gagaaagtccaanaattgtagaagaagaanaattgtagagtagaagaanaaagttaagaagaccgc	1440
Db	1381	GAGAAAGTCAAAATTTGAGAAAGAAATTTGAGAGTCAACAAAGAAAGTTATCCAAACCGC	1440
OY	1441	agcaaaagcttcaaacgatactactaaagaataagaaacgtctgaacgagatctatgat	1500
Db	1441	AGCAAAAGTCTAAACGATATCACTAAAGAGTATGAAGAAACTCTGACGAGATCTATGAT	1500
OY	1501	tccaaatccaacaataacacatcgaccctgacccaacttcgagaanaatgaatgagaaacggtac	1560
Db	1501	TCCAAATTCACCAATATACATCGACCTGACCAACTTCGAGAAATGATGGGAAACGGTAC	1560
OY	1561	tcttaagaattgtagaagaacgagacacccaataactctgatacctatagagaattctcaag	1620
Db	1561	TCTTACAAAGGAGGAACATGACACACCAATPAATACCTTGGCACTTATAGAAATTCAC	1620
OY	1621	cataactctgagaagcttcacacaagctcttaagataatgaaagactatctctcgtcgagac	1680
Db	1621	CATATCTTGAGAAAGCTCACCAAGCTTTAAGTATGAGAGACTATTCTCTCGGAC	1680
OY	1681	atgtgttgtgagaaagaacataaagtatatacaagaatctcaatgaagtacgcgaacagag	1740
Db	1681	ATTGTGTGGGAAAGAAAGAACTAAAGTATTACAGAAATCTCATAGTAAAGATGAGAAACGAG	1740
OY	1741	atcgagacgcgtgtgtggaacatlaagaagatgaaagacatgttttgagaagaagatt	1800
Db	1741	ATCGAGACGCTTGTGTGGAACATTTAAGAAAGATGAGAAACACTGTTTGTGAGAGAGATT	1800
OY	1801	acaaaagacganaaataaaccagatacgagaagatccctggaagctccgatattgttaaaatc	1860
Db	1801	ACAAAAGCGAAATATAACAGATGAGAAAGATCCTGAGAGTCTCGGATATTGTTAAAGTC	1860
OY	1861	caagtgcgaagatggtcctcctcatgaaacaagattgataaactcaagaagactcaactcat	1920
Db	1861	CAAGTGCAGAAAGGTGCTCTCATGAAACAAGATTGATGAACTCAAGACACTCAACTATT	1920
OY	1921	ctgagaagcgttgtagatttaaaaacataataacatgtgacgaataagttataagcagagagat	1980
Db	1921	CTGAAAGAAACGTGGATTAATAACATAATATACATGTGCCAAATAGTTATPAACAGAGAGAT	1980
OY	1981	aagcaggaacacatactacacccatcgtaaccgaagaagagataagcaaacctgaagtgctc	2040
Db	1981	AAGCAGGAAACCATPACTCTCATCTGTAACTAAGAAAGAGATTAAGCAAACTGAAGTGTTC	2040
OY	2041	atgcccacaaatcgagagcctgataccaacgagaagaagaagaacatataaactgaaagacag	2100
Db	2041	ATGCCCAAAATCGAGACCTGTGATCAACGAAAGAGAGAAAGAACATTAATAATCGAAGACAG	2100
OY	2101	tcaatataactccgagccttccaagaagaagaagataaaccggaacaggtctaccaccaagccc	2160
Db	2101	TCAATATACTCTCGAGCCTTCTCACAGAAAGAGAGATTAACCGGACAGCTACACCAAGGCC	2160
OY	2161	ggacaaacagcgcgttcaagctctcgaagacgataagctgccaagctcaagcacaagaagag	2220
Db	2161	GGACAAACAGCCCGTGTAGCTCTCTGAAAGCGATAGCGTTCAAAGCTCAAGCCACAAAGACAG	2220
OY	2221	aagaagcacaagcctccagatgycacagtgcaccttccaagaagctcaaaagctcaagtgctcata	2280

Db	2221	AAGCAGGACAGCCTCCACAGTGCACAGTGCCTCCGTTCCAGAGGCTTAAAGCTCAAGTGCCTACCA	2280
QY	2281	ccaccagctctcttgyaaataacaagaccggaanaatgttaagaaacttgactacttgaagaag	2340
Db	2281	CCACCAGCTCTCTGTGAATTAACAAGACCGGAAGATGTGAGAAACTGGACACTACCTTAGAAG	2340
QY	2341	ctctatgagctcccyaaataactccatcatctgcacaaataatactctcgtctctcaagc	2400
Db	2341	CTCTATGAGTTCCTGAATTAACCTCTACATCTGGCACAAATATATCTCTGTCCTCACAGC	2400
QY	2401	actatgaa-cgagaa-gattcttaaacagttacaagaataacaa-ggaagagagataactg	2460
Db	2401	ACTATGAACGAGGAAGATTCTTAAACAGTACAAATPAACCAAGGAAGAGAGAGTAAACTG	2460
QY	2461	tcctcttgytaagcaagagccgcgttgttaataataccaagaacaaatccgtatgfat	2520
Db	2461	TCCTCTTGTGTATCCACGTGGACCTGCTGTTCATTAATPCCAAACAAACATPCCGTATGTAT	2520
QY	2521	tcctatgctcgaataagccccaacaaatctctctccaaactgtcatgagaataataagaag	2580
Db	2521	TCTATGTTCCGATTAACCTCAACAAATTCCTCTCTCAACTGTTCAATGAGATATATAGAAg	2580
QY	2581	gagatggtctgcaaccctgtaataactaaagacaacgacaagattaa-gaaccctctgag	2640
Db	2581	GAGATGGTCTCCAAACCTGTATTAACCAAGAACAAAGCAACAAATTAAGAACCTTCTGAGg	2640
QY	2641	gaagctaa-gaag-gctcccaactctgtttaaactctctctccaagctctctcgaatcgaatacctg	2700
Db	2641	GAACTTAAGAGGCTCCACCCTCTGTTAAACCTCTCTTCACGCTCCATTCGAACCACTGg	2700
QY	2701	tcctccaacaccttaagaacaagcccggaagtga-gcttaacgacgaacacctcaaccgcgc	2760
Db	2701	TCTCTCAACCTCAAGAACGCCGAAGTGAAGGCTTAAGAGACACCTCTACTCGAC	2760
QY	2761	aaaccttaactcaactgaaactglttgagaacatccctgtctctcgcgagaagataagaac	2820
Db	2761	AACCTTAATAACTACTCAAGAAACGTTTGAACATCCTCTCTCGCGAAGATTAAGAAC	2820
QY	2821	atctaacaaagaacttattgagacagaatctgcgagaactcttaagaaagatctgaa	2880
Db	2821	ATCTPACAGAGACTTATTGGAACAAGAAATCGTCGGAACCTTCTACGAGAAGTACTGTAA	2880
QY	2881	gaagcgacacactctataacgagagcttaacaaactctgttaaatctaaagccgatat	2940
Db	2881	GACAGCGACAAATTCTTAACGAGAGCTTCACTAATTGCTGAATCTTAAGCCGATAT	2940
QY	2941	atcaactctcttaa-cgatactaa-cgtaag-cgtaag-cgtaag-agatactaaactg	3000
Db	2941	ATCAACTCTCTTAACGATGAATCTTAAACGTAACAACTCTGGAAGAGACATCAATTAAGCTG	3000
QY	3001	aagaagacactcgcaactcgaaactctcgaaactgtatacaaaatatacaacttgaacttgaaga	3060
Db	3001	AAGAAGACACTGCAACCTGAACTTCGACCTGTATCAACAACTGAACCTGAACCTGGAGGA	3060
QY	3061	ctcttcgacaagaagagacagctccgcgaagtataagaatgcagatacaagaagtctgaactg	3120
Db	3061	CTCTTCGACAAAGAAAGACAGCTCGGCAGATATAGATCCAAATCAAGAAGTTCACTTG	3120
QY	3121	ctcaagagacagcttgaagaacaaactcaactcacttgaacaaatccgaaacacgtactgag	3180
Db	3121	CTCAAGGAGCGAGTTTGAACCAAACTCACTCACTGAACAACTCCGAACACGTAAGTACGAG	3180
QY	3181	aactctcagatgtctctcaacaagaagaaggaagccgagatacgccgagacagagaacact	3240
Db	3181	AACCTTCTCAGTTCCTTCAACAAGAAAGAAAGAAAGCCGAATGCGCGAACAAGAAACT	3240
QY	3241	ctggagaacaccagaaattctctccaacacttaaaagagctcgctcaaglatataaagtgc	3300
Db	3241	CTGAGAAACCAAGATTCTTCTCAACACTTCAAAGAGCCTCTGTAAGTATTATATATGCG	3300
QY	3301	gagctctctctctgaaactctctccgagagagatcaatcaagcagagataactaagcc	3360

Db 3301 GAGTCTTCTCTGTAAGACTCTCTCCGAGAGAGACTCAGACCGAGATCAACGCC 3360
 Qy 3361 agccctcggaacttcaagtcctgtcttaagctcgaaggaactgaagcaactgaaac 3420
 Db 3361 AGCCCTGAGAACTTCAAGGCTCTGTCTAAGCTGGAAGGCAAGCTGGAAGCAACCTGAAC 3420
 Qy 3421 ctggaagaagaagaactcagctacactctccttaagcgagctcaccactgtctcgagctc 3480
 Db 3421 CTGGAAGAAGAAGAACTCAGTACCTCTCTAGCGGAGCTGCATCACCCTGATGCCCGAGCTC 3480
 Qy 3481 aaggaatcatttaagaacaagaactacaccggcaatagcccaagcgagaaataatacagac 3540
 Db 3481 AAGGAAGTCAATTAAAGAACAGAACTACACCGCAATAGCCCAAGCGAGATATATACAGAC 3540
 Qy 3541 gtgaataacgcacctggaactcttacaagaagttccctcgccgaaggaagcaatgtcccat 3600
 Db 3541 GTGAATTAACGCACTGGAACTCTTACAGAAGTTCTGCTGGAAGGAACCAATGTGCGCACT 3600
 Qy 3601 gtgtgtctgaaactgtgtgtccgacacactgtagcagttcctaactaagaagctgtcatct 3660
 Db 3601 GTGTGTCTGAATCTGGCTCCGACACACTGAGAGAGTCCAACTAAGAAGCCTGATCT 3660
 Qy 3661 actcattgtcgagagccgagtgccataataataccacatactcagaacgttcgaatgaagtc 3720
 Db 3661 ACTCATGTGCGAGCCGAGTCCCAATACATTTACACATCTCAGAAAGTCCAGATGAGGTC 3720
 Qy 3721 gatgacgtcatcatgtgtgtccctcttcgagagagcggaaggaactgaatgactcggc 3780
 Db 3721 GATGACGTCACTATGTGTGCTTCTTCCGCGAGAGCGAGAGAGACTAGCATACCTCGGC 3780
 Qy 3781 caggtgtgtccacccgtgtgaggtgtgtacactcccttcggtatgtataaatctgtccaaatc 3840
 Db 3781 CAGGTGTGTACCCGAGAGGCTGTCACTCCTCCGATGTATACATTTGTGCCAAATC 3840
 Qy 3841 gagaaagataaagaagtgctctctatctgaagaaactctgagcgagtgctatagtgctcgaag 3900
 Db 3841 GAGAAAGAAATAGGAAGTCTATCTGAAACCTCTGCAAGGCGTCTATGTGCTCTCAAG 3900
 Qy 3901 aaacagctggagaataaagaactgtgaactcctcaatgtccaagtgtaaggaactctgaacagc 3960
 Db 3901 AAACAGCTGGAGAAATACCTGATGACCTTCACTGTCACAGCTAGAGACCTTTGAAACAGC 3960
 Qy 3961 cgctttaaagaagaagaacttcaagaacgtctgtgagagcgacttgatctccctataaa 4020
 Db 3961 CGCTTTAATAAGAGAGAAATTTCAAGAACGCTTGAGAGCGACTTGATCTCCATATAA 4020
 Qy 4021 gacccgacactccttaactcaagtgctgaagagccatacaagttcctaataaagaag 4080
 Db 4021 GACCTGACCTCTCTTAACCTGCTTCAAGGACCCATACAAAGTCTCAATATAAGAGAG 4080
 Qy 4081 aggaataaattctgtctagttacaactataatacaagactccatcgacacccgatatcaat 4140
 Db 4081 AGGATAAATTTCTGTCTAGTTACAACTATATCAAGAGCTCATGACACCCGATATCAT 4140
 Qy 4141 ttgcgttaatgctgtcgtgggtattatacaagatcctgagcgaaataatacaagtctgacct 4200
 Db 4141 TTGCGTAATGANTGCTGGGGATTATCAAGATCTGAGCGAAATAATACAAAGCTGACCTT 4200
 Qy 4201 gactctattaaaagaatatacaagaataagcaagcggaagaaataatactgtcccttc 4260
 Db 4201 GACTCTATTAAAAAATATATACAGATTAAGAGCGAGAAATGAAAAAATATGTGCCCTTC 4260
 Qy 4261 ctgaataacatcgaaacactgtacagaacagtgaaacgaacaaatgacctcttcgtaatt 4320
 Db 4261 CTGAATTAACATCGAAACCTGTACAGAGAGTGAACGACAAATGACCTCTTCTGTAATT 4320
 Qy 4321 caactgtgagcgcaaggtccctcaactatacttaagaaagcaatgtgtgaagttaaactc 4380
 Db 4321 CACCTGAGAGCGCAAGGCTCTCAACTATACTTACGAGAAGAGCAATGTGGAAGTTAAATC 4380
 Qy 4381 aagaggtcgaactactcaaaaaacatcaagaagcaagctgtgcgagattcaagaaaaaataac 4440
 Db 4381 AAGGAGCTGAACCTACCTCAAAAACATCCAAAGCAAGCTGCGAGATTTCAGAAAAATAAAC 4440

Qy 4441 aatttcgtcggaattgcgaacactgtctaccgatatlaaccacaacaaatctctgaccaa 4500
 Db 4441 AATTTCGTGGAATTGCAGACTGTCTTACCGATTATTAACCAACAATCTCTTGACCAAG 4500
 Qy 4501 ttctgtccactgcatgggtgttcgaaaaaccccgccaaacagttctgagcaatctgtc 4560
 Db 4501 TTCTGTCTCAGTGGCATGGTGTTCGAAACCTCGCCAAACAGTGTGAGCAATCTGCTC 4560
 Qy 4561 gacggcaacctgcagggcatctgtaacatctcccaagcccaatgtgtgaagaacagtc 4620
 Db 4561 GACGGCAACCTCGACGGCATGTGTGAACATCTCCACAGCAACCAATGGCTGAAGAAAGTGC 4620
 Qy 4621 ccccaagaatgcgctgtgttcagagcatctgagcgagcgcggaagtgcaagtgctcc 4680
 Db 4621 CCCCAAGATAGCGCGCTGTTTTCAGGCACTTGAGCGAGCGGAAAGATGCAAGTCTCTG 4680
 Qy 4681 aactacaacaagaagaagataagtgctgtgagaaccccaacccctacactgcaatgaaac 4740
 Db 4681 AACTACAACAAGAGAGATAGTGCCTGAGAACCCAAACCTTACTGCAATGAAAAAC 4740
 Qy 4741 aatggcggtgtgtgacgccgagttaaatgtcacccggaggaagacaggtcttaacggaag 4800
 Db 4741 AATGGCGGCTGTGACGCGCATCTAAATGCACCGAGGAAGACGCGCTTAACGGAAG 4800
 Qy 4801 aaatcacatgtgagtgtaactaagcccgactctataccactcttcgaaggaatttttc 4860
 Db 4801 AAATCACAATGAGAGTGTACTAAGCCCGACCTCTTATCCACTTTTGACGAGATTTTTC 4860
 Qy 4861 tccagctcaattctccctgggaactcctctcctgtcatctcaatgactgtacactgacgc 4920
 Db 4861 TCCAGCTCAATTTCTCGGCACTCTCTTCCGTGATCTGATCCTGATCTGTACAGC 4920
 Qy 4921 ttcatctaataagatcgaag 4940
 Db 4921 TTCATCTAATAGATGATG 4940

RESULT 2
 LOCUS CVEI31294 4940 bp DNA SYN 24-FEB-1999
 DEFINITION Cloning vector pBSK*-MSP-1s/FCB.
 ACCESSION AJ31294
 VERSION AJ31294.1 GI:4455908
 KEYWORDS Cloning vector; major surface protein-1; msp-1 gene.
 SOURCE Cloning vector pBSK*-MSP-1s/FCB.
 ORGANISM Artificial sequence: vectors.
 REFERENCE 1 (bases 1 to 4940)
 AUTHORS Bujard, H.
 TITLE Direct Submission
 JOURNAL Submitted (14-DEC-1998) Bujard H, ZMBH - Zentrum fuer Molekulare Biologie, Universitaet Heidelberg, Im Neuenheimer Feld 282, 69120 Heidelberg, BW, GERMANY
 REFERENCE 2 (bases 1 to 4940)
 AUTHORS Pan, W., Rayov, E., Tolle, R., Frank, R., Mosbach, R., Turbachova, I. and Bujard, H.
 TITLE Vaccine candidate MSP-1 from Plasmodium falciparum: a redesigned full-length protein from Escherichia coli and mammalian cells
 JOURNAL Nucleic Acids Res. 27 (4), 1094-1103 (1999)
 MEDLINE 99128299
 FEATURES
 source location/Qualifiers
 1..4940
 /organism="Cloning vector pBSK*-MSP-1s/FCB"
 /db_xref="taxon:85834"
 /focus
 1..4940
 /organism="Plasmodium falciparum"
 /strain="FCB-1"
 /db_xref="taxon:5833"
 10..4929
 /gene="msp-1"

```
sig_peptide      10..69
                  /gene="msp-1"
                  /gene="msp-1"
                  /function="surface protein"
                  /evidence="experimental"
                  /product="major surface protein-1"
                  /protein_id="CAB36903.1"
                  /db_xref="gi:4455909"
                  /translation="MKITFLCSEFLPIINTOCVTHSEYQELVKKLEDAEDAVLTGYS
                  LFOKRWVNLSESTGAVTSTPGSGSVASGSGSVASGSGSVASGSGSVASGSGSVAS
                  GSGSRRTNPSDSDAKSYADLKRNVNLTLEKLPQLEFDLTNNHLLTCDN
                  IHGKYLIDGEEINELLYKLNFPDLRAKINDYCANCOIPNNLKIRAMELVYK
                  KLVFGKRPDLINKDNGKMEYIKKNNKTENINELIEESKPTDKNKNAKKEEK
                  KLYOAYDLSTNNKOLEAHNLSVLEKRIIDLIKNNELDKDINEKLNKPPANS
                  NTPNLLDKNNKIEBEHEKIKTAKTIFKNIDSLPTDLELEYIIREKKNIDISAKV
                  ETKESTEPENYNGVTPPLSYNDINNALNELNSFDLI.NPFDYTEPSKNITTDNRK
                  KEINIEKRIKIEKKIESDKSYEDRSKSLNDITKEKLEINIEYDSFNNNIDLTN
                  FEKMGKRSYKVEKLTHTNPFASYENSHNLEKLTALKAYMEDSLRNIVVEKELKY
                  YKNLSKIEINEETLVENIKDEOLPEKTIKDNKPKETLLEVDIYKVOYOKVL
                  MNKIDELKKTOLILKNELKKNHYPNSKQENKQEPYLIYLKKEIDKLYFMKVE
                  SLINEKKNITQESDSESTGEITGQATTKQOQAGSLEDSDVOAQOEKQA
                  QPVPVPVEAKAQTTPPAVNNKTEVNSKIDYLEKLEFNLNTSYICHKYLVSHT
                  MNEKILKQYKITEESKLSGCDPLDLFNIONNIPVMSMPSDLSNLSOLFMEITE
                  KEMVGNLYKKNDKIKNLLEAKKYSTVSTLSSSQMPLSLTPQDKEVANDST
                  HSTNNLSKLLENILSLGKNKIYQELIGOKSSNPEYKILKDSDFNESFTNFK
                  SKADINDLNDESKRRKLEEDINLKLTLOSLFDLYNKKIKLEKLPDKKTVGYTKM
                  QIKLTLLLEOLSKLNSLNKPHVLQNSVFNKKKELEIAETNLTENTLILKLY
                  GILVYVNGESSPLKTLSEESIQEDNVALENFVLSLEKLDNINLEKKLSYL
                  SSGHLHLAELEKVINKNNTGNSPSENNVDNNALESYKFLPGTDVAATVVSQS
                  PLEOSOPKRPASTHVGAESNTTTSOENVDEVDVITVPIGSEEDYDLDGOVYTG
                  EAVMPSVIDNLISKIENEYEVILYKPLAGVYSILKOLENNMTFNVMYKDLINSRN
                  KRENKRNLESOLIPKDLTSSNIVYKPKYKFLAKKRRKPI.SSTNVIKDSIDTDINF
                  ANDVLGTYKILSEKTSKSDLSIKTYINRQGENEYLLPPLNIEFLYKTVNDKIDLV
                  IHLKRVLNTEYIEKSNVEYKIKELNYLKTIDKLADPKNNNFV.IADLSITNNHL
                  LTKFLSTGMVENFLAKTVLSNLDGLOGLMISOHQCKKCPGNSFRILDBREE
                  CKCLNTKQEGDKVEVNPNCNENNGNDADAKCTEEDSGSNGKIKTCECTKPPSY
                  LFDGIFGSSSNLGIFFLLIMLILYSFI"
```

BASE COUNT 1669 a 1156 c 1060 g 1055 t
ORIGIN

Query Match 100.0% Score 4938.4; DB 56; Length 4940;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4939; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
1 cgcacgcgtatgaaatcattcttcctctgttcattctctgttttcaatcaact 60
1 CGCAGCGGTATGAATTCATTTCTCTCTGTCATTTCTGTTTATCATCATACT 60
61 cagtcgtgacccacgaatccatcagaagcgtgttaagaactggaagccttgaagat 120
61 CAGTGTGACCCACGAATCTATCAGGAGCTGCTTAAAGAACTGCAACCTTTGGAAGAT 120
121 gccgctctaccggaataagccgtgttcacagaaggaagatggttgaataaaggagc 180
121 GCCGCTCTTACCGGATACAGCCTGTTCCAGAAAGAGATGCTGTGATGAAGGAGCG 180
181 agtggcagcgcdttaaacaacagcacccggtttcctaaaggctctgtgctaaagcgg 240
181 AGTGGCAGCGCGGTAAACAACAGCACCCGGTCTTAAAGGCTCTGTGCTACCGGTGCG 240
241 tccggttggtctgtggtctctcgtggtctcgcctccggcgagcagcgtgacataagt 300
241 TCCGGTGGTCTGTGCTCTCTGGGGGCTCTGTCGCTCCGGCGGACGCTGGCATAGGT 300
301 agcctcagtgcaagcggcggtctccgggaacagctcgaagaacaaatccatctgaact 360
301 AGCCTCAGTGCAAGCGGCGGTCTCCGGGAACAGCTCGAAGAACAATCATCTGCAACTCT 360
361 agcagattccgacgcaagctctcagcgaactcaagcacagagtggaactatctccctc 420
|||||
```

```
Db 361 AGCAATTCGACGCCAAGTCTTACGCCGACCTTCAGCACCGAGTGAANAATCTCTCTC 420
421 actataaagagctgaagtaagtaagtaagtaagtaagtaagtaagtaagtaagtaag 480
421 ACTATAAGAGCTGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 480
481 tggataaacattatgcttcaataatacgtatgtaagcgttgaagaagatacaatgaact 540
481 TGGATAAACATTATGCTTCAATAATATGCTGATGAGAGGTTAGCAAGATCAATGAACTC 540
541 ctgtacaagtgtaattcttactctgactctgactaaggccaactgaatgaactgtcgcc 600
541 CTGTACAAGTGTGAATTTACTTCACTTGCATTCGCTAAGGGCCAAACTGAAATGACGTTT 600
601 aagactattgtcaaatccatcattcaattgaagaataagccaagatgtgaagcgtatgt 660
601 AAGACTATTGTCAAAATCCATTCAATTCGAATGAGACGCCAACGAGTGGACGTAATG 660
661 aagaagctgcttcgagatacgaagcctctcgacaacatacaaggaacatgtgggaag 720
661 AAGAAGTGTGCTTCGATATCGCAAGCCTCTGCACAACTCAAGACATATGCGGAAG 720
721 atggaagattataataaagaataagaagaccatcgagaacataaagcgtatcgaa 780
721 ATGGAAGATTATATTAATAAAGAAATGAAGACCATCGAAGCATTAACGAGCTGATCGAA 780
781 gaatcaaaaaagccatagaacaaataagaatgcaacccaaggaaggaagaaagaaag 840
781 GAATCAAAAAAGCCATTAGCAAAATAAGAAATGAAGCAACCAAGGAGGAGAAAGAAAG 840
841 ttgtacaaagccagtagcagcctctcactatatacaaaagcttgaagaagccataac 900
841 TTGTACAAAGCCAGTAGCAGCCTCTCATTCTATTAACAAACAGCTTGAAGAGCCCTAAC 900
901 ctcaatcagctactgtagaagcgcatalagacacccctcaagaagaatagaataatcaaga 960
901 CTCAATCAGCTACTGTAAGAGCCGATAGACACCCCTCAAGAAATGAATAATCAAGAA 960
961 ctgctcgaacagaataatgaatgaatcctccgacgaagcgaactctggaagacccct 1020
961 CTGCTCGAACAGATTAATGAATTAAGAAATCCGCCACGCCAACCTCGGAGACACCCCT 1020
1021 aacacgctcgtgacagaagaagaagaatagagaagcagaagaagaagaatcaagaagac 1080
1021 AACACGCTCTGTCAGACAAGACAAGAAATAGAGAGACACGAGAAAGATCAAGAGATC 1080
1081 gccaaaaccattaaagtccaacatagatctctcttactgataccctctgagctgagatc 1140
1081 GCCAAAACCATTAAGTTCAAACATAGATTCCTCTTACTGATCCCTTGAGCTGGAGTAC 1140
1141 tacttgaagagaagaataagaatataagacatctccgcaaaagctgagacaaagaatca 1200
1141 TACTTGAAGAGAGAATTAAGAAATATAGACATCTCCGCAAAATCGAGCAAAAGAAATCA 1200
1201 accgaacctaataatgaatatccaaatggtgacgtaccctctgtcttataacgatatcaac 1260
1201 ACCGAACCTAATGAATATCCAAATGCTGACGTACCCTCTGTCTTATTAACGATATCAAC 1260
1261 aacgctctcaagcagctcaatagcttcggttgacttgatlaaacccctctgattatacga 1320
1261 AACGCTCTCAAGCAGCTCAATAGCTTCGGTGACTTGATTAACCCCTTCGATTATAGAAA 1320
1321 gaaccccttaagaatacttaacacagaacatgagagaagaagaatctcaacgaatacaag 1380
1321 GAACCCCTTAAGAAATATCTTAACAGACAAATGAGAGAAAAGTTATCAACGAAAACAAG 1380
1381 gagaagaatacaaatgagaagaagaagaatgagagtgacaagaagaagaatgaagaagcgc 1440
1381 GAGAAGAATACAAAATGAGAAGAAGAAGAATGAGAGTGAACAAGAAAAGTTATCAAGAGCC 1440
1441 agcaaaagcttaaaagatactcaataaagatgaagaagctgctgaagaagatactatga 1500
1441 AGCAAAAGCTTAAAGATATCACTAAAGAGTATGAAGAACTGCTGAACGAGATCTATGAT 1500
```


QY 1501 tccaaatccaacataacalcgcacctccgaacacttcgagaaaatgtagggaaaaacggttac 1560
|||||
Db 1501 TCCAAATTCACACATATACATCGACCTGACCACATTGAGAAAATGATGGAAAACGGTAC 1560
QY 1561 tcttacaagctggagaacatgacacacccaataactctgtacatcctatgaaatctaaag 1620
|||||
Db 1561 TCTTACAAGGTGAGAACTGACACACCAATTAATACCTTGACATCTGATGAAATTCGTAAG 1620
QY 1621 cctaatcttgagagaagctccccaagctcttaagtatatggagacattctctggggaac 1680
|||||
Db 1621 CATAACTTGAGAGCTCACCAAGCTTTAACTATATGAGAGCAATTCCTGCGGAAC 1680
QY 1681 atgtgttgagagaagaacataaagatlaacaagaatcctaaagtaagatcgaagaacgag 1740
|||||
Db 1681 ATGTGTTGGAGAAAGACTAAAGTATTCAGAAATCTCATAGTAAAGTGCAGAAACGAG 1740
QY 1741 atcgagaagctgtgtgagaacatlaagaagatgaaagaaagctgtgttgagaagaaatt 1800
|||||
Db 1741 ATCGAGACCTGTTGTGAGAACATTAAAGAGATGAAGAACAGTTGTTGAGAGAGAAATT 1800
QY 1801 acaaaagagaataataacacagatgagaaagatcctggaagctccgaattgttaaagtc 1860
|||||
Db 1801 ACAAAGAGAGAAATTAACCAAGATGAGAAAGATCTGAGGCTCCGATATGTTAAAGTC 1860
QY 1861 caagctgagaaggtgtcctcctcaatgaaacaaagatgtatgaactcgaagaactcaactcat 1920
|||||
Db 1861 CAAGTCAGAGAGGTGCTCTCCTCATGAACAGATGTGATGAAGTCAAGAAAGTCAACATCATT 1920
QY 1921 ctgagaagaagctgtgaaatlaaacaataatacatatgtgtccgaatgttaagaagaaat 1980
|||||
Db 1921 CTGAGAAAGCTGTGAGTTAAACATTAATATACATGTCCGATAGTTATTAAGCGAGAAAT 1980
QY 1981 aagcagaagaacactactactactcactcactcaagaagaagatagaacaactgaaagcttc 2040
|||||
Db 1981 AAGCAGGAACCATACTACTACTCTCATCTGACTCAAGAAAGATGACAACTGAAAGTGTTC 2040
QY 2041 atgcccagaagctgagaagccttgatcaacgaagaagaagaacatlaaactgaaagagacag 2100
|||||
Db 2041 ATGCCCAAGGTGAGAGCCTGATCAACGAAGAAAGAACACTTAAGTGAAGACAG 2100
QY 2101 tcaagataactccgaagccttcacacagaagaagataccggagacagctccacaaagccc 2160
|||||
Db 2101 TCAGATAACTCCGAGCCTTCACAGAAAGAGATTAACCGAGACGCTACCAACCAAGCCCC 2160
QY 2161 ggaacaacagagcggttcagctcctcgaagcgatagcgtgcaagctaaagcaacaagagcag 2220
|||||
Db 2161 GGACAAACAGGCGGTTCACTCTCGAAGCGGATAGCGTGCAAGCTCAAGCACAGAGCAG 2220
QY 2221 aagcagagcagagcctccagtgccagtgccgttccagaggtctaaagctcaagtgcttaca 2280
|||||
Db 2221 AAGCAGAGCAGAGCTCCAGTGCAGAGTGCCTGTCAGAGAGCTTAAGCTCAAGTCCATCA 2280
QY 2281 ccaacagagcctcgtgtgaaatacaagaagcagaatgtcagcaaacctgtaactgaaag 2340
|||||
Db 2281 CCACCAAGCTCTGTGTAATTAACAAGACCGAAGATGTAGCAAACTGAGCTACCTTGGAAG 2340
QY 2341 cctcatgagcttctgaaatacatcctcatcatctgcaccaaaatatacctgtctctcaagc 2400
|||||
Db 2341 CTCTATGATTTCTCTGATACATCTCAATCTGCACAAAATATATCTCTGCTTCACAGC 2400
QY 2401 actatgaagaagagaattctlaaacaagtaacaagataccaagaagaagaagagtaaaccty 2460
|||||
Db 2401 ACTATGAAGAGAGAAATTTCTTAAACAGTACAAAGATTAACCAAGAGAGAGAGATTAACCTG 2460
QY 2461 tccctctgtgatccactgagacctgtgttcaatataccagaacaacatactccgttlatgat 2520
|||||
Db 2461 TCCCTCTGTGATCCACTGAGCCGTGTTCAATATTCAGAACAACTATCCCGTTATGTAAT 2520
QY 2521 tctatgttctgataagcttcaacaattctctctctcaacgttctaatgtatgagatataatgaaag 2580
|||||
Db 2521 TCTATGTTGATAGCTCTCAACAATTTCTCTCAACAGTGTATGAGATATATAGGAAG 2580
2581 gagaatgtctgaacacctgtataaactcaagaagacagacaagaatlaagaacctctggag 2640
|||||
Db 2581 GAGATGTTGCGACACTGTATTAACCTCAAAAGAACAGACAGATTTAAGAACCTTTCGAG 2640
QY 2641 gaagtaagaaggtctccacctgttlaaaactctctctccagctccatgcaaccactgt 2700
|||||
Db 2641 GAAGCTTAAGAGAGGTCTCCACCTCTGTTAAACCTCTCTCCAGCTCCATGCAACACAG 2700
QY 2701 tctctcaaccttaagaagaagcccggaagtgagagctcaagaagaacacctctcatagac 2760
|||||
Db 2701 TCTCTCAACCTTAAGACAAAGCCGAGTGAAGGCTTAAGACACACACCTCTCACTGACCC 2760
QY 2761 aaccttaataactcaactgaaactgttctgaaacatcctgtctctcggcaagaataaagac 2820
|||||
Db 2761 AACCTTAATAACTCACTGAACAGTGTGAGAACATCTCTCTCTCGGCAAGATAAGAAC 2820
QY 2821 atctacaagaactatctggaagaagaatcgctcgagaactcttagagaagaatactgaaa 2880
|||||
Db 2821 ACTTACCAAGAACTTATTTGAGAGAAATGCTCCGAGAACTTCTACGAGAAATACTGAAA 2880
QY 2881 gacagagacaactctcatacgaagagctcactaactctgtgaaatctaaagccgaatg 2940
|||||
Db 2881 GACAGCGACACATTTCTATTAACGAGACTTCACTTACTGTGTGAATCTAAAGCCGATGAT 2940
QY 2941 atcaactctcttaacgaatgaatlaaagcgtlaagaagctggaagaagaca tcaataagctg 3000
|||||
Db 2941 ATCAACTCTCTTAACGATGAATCTAAACGTAAAGAACTGGAAGAGAGCAATCATATAAGCTG 3000
QY 3001 aagaagaacactgcaactgagctctggaactgtacacaagttacaactgaaactgtgagaga 3060
|||||
Db 3001 AAGAAACACTGTCAACTGAGCTTGACCTGATCAACAACTTCAAACTGAACCTGAACTGAGAA 3060
QY 3061 cctctcgaagaagaagaagcagctcggaagataagatgcagatcagaatgtgaactgtc 3120
|||||
Db 3061 CTCTTGACAAAGAAAGAGACAGTGGCCAAAGTATTAATGCAATGCAAGATTTGACTCTG 3120
QY 3121 ctcaagagagcagcttgaagaacaaactcaactcaactgaacatccgaagaacagctactgcag 3180
|||||
Db 3121 CTCGAAGAGAGAGCTTGAAGCAAACTCAACTCACTGAACATCCGAACACGTAAGTGCAG 3180
QY 3181 aactctcagctgtctctcaacaagaagaagaagccggaatctgcggaagcagaagaact 3240
|||||
Db 3181 AACTCTTCAGTGTCTTCAACAAGAAAGAAAGGAAAGCCGGAATGCCCAAGACAGGAACACT 3240
QY 3241 ctggaagaacacagaattctctcaacaactcaaaagagctctgcaaatatlaaagtc 3300
|||||
Db 3241 CTGGAAGAACCACAGATTTCTTCAAAACACTCAAAAGGCTCTGTCAAGTATTTATATGGC 3300
QY 3301 gagctctctctctgaagactctctcgaagagagcatccagaacggaataactagccc 3360
|||||
Db 3301 GAGCTTTCCTCTGAAAGCTCTCTCGAGGAGAGCATCCAGACGAGATTAACCTACGCC 3360
QY 3361 agcctcgaagaacttcaaggtctctgtctaaagctcgaagcgaagctgaaagcaacctgaaac 3420
|||||
Db 3361 AGCCTCGAAGAACTTCAAGGTCTGTAAAGCTCGAAGGCAAGCTGAAGACAAACCACTGAAC 3420
QY 3421 ctggaagaagaagaagcttagagctactctcttagcgagactcatcaactgtatccggagctc 3480
|||||
Db 3421 CTGGAAGAAGAAGAGCTCACTACCTCTTAGCGGAGCTGCATACCTGTATGCCGAGCTC 3480
QY 3481 aaggaagctatlaagaagaagaactcacccgggcaatagaccgaagcgaagataataaagac 3540
|||||
Db 3481 AAGGAAGCTATTAAGAACAAAGAACTACACCGGCAATAGGCCAAAGCAGATATATACAGAC 3540
QY 3541 gtgaataacgcaactggaatcttaagaagaatctctgtcctgaaaggaacagatgtcgccact 3600
|||||
Db 3541 GTGAATTAACGCACTGGAATCTTACAAAGAGTTCCTCCGGAAGAAACAGATGTCCGCACT 3600
QY 3601 gtggtgtctggaatctgtgctcggaacacagctgagagagcttccaacttaagaagctgtcatct 3660
|||||
Db 3601 GTGGTGTCTGAATCTGGCTCGACACACTGTGAGCACTTCAACCTTAAGAAAGCTGTGATTT 3660
QY 3661 actcatgtctggagcgagctcaataataactcctcagaagctgcagatgaggtc 3720

Db	3661	ACTATGTCGAGGCCGAGTCCAAATACAAATTACCAATCTCAGAACGTGCACGATAGGTC	37200
Oy	3721	gatgacgtcatalctgttgcctatcttcgggagagagagagactacagatbaacctgcgc	37800
Db	3721	GATGACGTCATCTTGTCCTATCTTCGGCGAGAGAGAGAGACTACGATACCTCGGC	37800
Oy	3781	caagtgtcacccggttgagcgttcgaactcctccgattgatgaatacattcgtccaaatc	38400
Db	3781	CAGGTGTCACCCGGGTGAGCGCTGCACCTCTTCGGATTGATTAACATTCTGCCAAATC	38400
Oy	3841	gagaaacgaatacgaatgacctatctgaaacctctggcagcgctataagtctctcaag	39000
Db	3841	GAGAAACGAATCGAAGGCTCTATCTGMAACCCTTGCGGGGTCTATAGGCTCTCAAG	39000
Oy	3901	aaacgcgcggagaaatacgcgatgacctcaatgctaaagtgaaagacatcttgaaagc	39600
Db	3901	AAACGCGTGGAGAAATTAACGCGATGACCTTCAATGCTAACGTGAAGCACTTCTGACAGC	39600
Oy	3961	cgctttaataagagagaaaaattcaagaacgctcttgagagcgactgtaattccataaa	40200
Db	3961	CGCTTTAATAGAGAGAAATTTCAAGAACGCTTGAGAGAGGAGCTGTATTCCTATATA	40200
Oy	4021	gacctgacctccttaactacgttgcgaaggccaatacaagttccataaagaagaag	40800
Db	4021	GACCTGACCTCTCTTAACCTAGTGTCAAGAGCCATCAAGTTCCCTCAATTAAGAGAG	40800
Oy	4081	aggagataattcgtcttaattcaacatataraaagatccatagaacccgaattcaat	41400
Db	4081	AGGGATAAATTTGTTCTAGCTACCACTATATCAAGAGCTCCATCGACCCGATATTCAT	41400
Oy	4141	ctgcataatgctgtgcgagggtatctcaagatccctagcgaaataatacaagctbaact	42000
Db	4141	TTGCGTATATGATGTGCTGGGGTATTACAAGATCCTGAGGAAAAATACAGTCTACCTT	42000
Oy	4201	gacctatataaagatatacaacgataagcaagcggagatgaaataatctgccttc	42600
Db	4201	GACCTATATAAAAGTATATCAACGATAGCAAGCGGAATAAATAATCTGCCCTTC	42600
Oy	4261	ctgataaacatcgaaaccccgctacaagaacggtgaagagaaaatcgacctctgtlaat	43200
Db	4261	CTGATATAACATCGAAACCCGTGTCAAGACGTGAACGAAGAAATGACCTCTTCTTAATT	43200
Oy	4321	caacctgagagccaaggtccctcaactactactcaagaagaagaatgtggaagttaaact	43800
Db	4321	CACCTGAGAGCGCAAGGTCCCTCACTTACTTACGAGAAAGCAATGTGGAATTAAAAATC	43800
Oy	4381	aaggagctgaaactacctcaaaacaatccaagaacgctggcagatctcaagaanaataac	44400
Db	4381	AAGGAGCTGAACCTCACTCAAAACAATCCAAGACAAGCTGCGAGATTTCAAGAAAAATAC	44400
Oy	4441	aatttcgtggaattgcaagacctgctacacgattataaccacaacaactctctgcaag	45000
Db	4441	AATTTCTGTGGAATTTGCAAGACCTGTCTACCGATTATTAACCAACAATCTCTTGACAAAG	45000
Oy	4501	ttctgtccacctgcatcggtgttcgaaaacctgccaanaagtcctgacaactctgtc	45600
Db	4501	TTTCTGTCCACTGGCATGGTGTTCGAAAACTTGCCAAAACGTGCTGACATTTGCTTC	45600
Oy	4561	gacggcaacctgcagggcagctgtaaacatctccagcgaccaaigtgcttgaagaacaagtc	46200
Db	4561	GACGGCAACCTGCAGGGCATGTGTAACATCTCCACACCAATATGCGTGAAGAACAGTGC	46200
Oy	4621	ccccagaatagcgcgctgtcttcaggaactcggacagcgcggaagatgcaagtgtctcctg	46800
Db	4621	CCCCAGATATACCGCTCTTTCAAGCATCTTGGAAGACCCGGAAGAGTGCMAAGTCTCTCTG	46800
Oy	4681	aacctacaacaagaagagataaagtgcgttgagagaaaccaaccctactctgcataaanaac	47400
Db	4681	AACCTACAACAAGAGAGATTAAAGTGTGCTGAGAACCCCAACCTTACTGCAATGAANAAC	47400
Oy	4741	aatggcggtgttgacgcgcgaatgctaaatgcaccgagaaagacaggtgcttaacgaaag	48000

[illegible]

SSGLHHLIAELKEVIRKNTGNSPENNNDVNNALLESYKFLPEGTDAIVSESGS
DLEQSOQRKPASTHVGAESNTITTSQVNDDEVDVITVPIGSESEDDDGOVYTG
EAVPSPVIDNITLSIKIENYEVLTKPLAGVRSIKOLENNMTFNVANKDILNRFN
KRENNKYLESDLIPIKDLTSSNYYKDPKFLNKKRKLPSSTNYIKDSIDTPIFN
ANDVGYIKIIESEKSKSDSDIKITNDQGENEYLFPLNNITELVKVNTDLEFV
IHLEKVLNNTYEKSNVEKIKELNYLTIDOKLADFKNNNEVIADLSTDYNNNL
LTKFISTGMFENLAKTVLSNLDGNLQGLMNI SOHCYKCKGCPNSGCFRLIDRE
CKCLINTYKOEQCEVNEPNPCENNGCGDADAKCTEEDSGSNGKKTCECTKPDSP
LEDGIFCSSNNLGISFLILMLILYSFI"

mat_peptide
/product="region coding for 83K, 42K and 19K major surface
antigens (aa 1-1628)"
631..728

repeat_region
/note="alternating repeated SVA-SGG (aa 72-104)"
763..765

misc_feature
/note="pot. glycosylation site"
1219..1221

misc_feature
/note="pot. glycosylation site"
2707..2709

misc_feature
/note="pot. glycosylation site"
2719..2721

misc_feature
/note="pot. glycosylation site"
2764..2766

misc_feature
/note="pot. glycosylation site"
2947..2949

misc_feature
/note="pot. glycosylation site"
3175..3177

misc_feature
/note="pot. glycosylation site"
3307..3309

misc_feature
/note="pot. glycosylation site"
3589..3591

misc_feature
/note="pot. glycosylation site"
3910..3912

misc_feature
/note="pot. glycosylation site"
3937..3939

misc_feature
/note="pot. glycosylation site"
4750..4752

misc_feature
/note="pot. glycosylation site"
636 c

BASE COUNT 2644 a 636 c 727 g 1910 t

Query Match 57.0%: Score 2816.4; DB 96; Length 5917;
Best Local Similarity 73.2%: Pred. No. 0;
Matches 3609; Conservative 0; Mismatches 1321; Indels 0; Gaps 0;

QY 10 atgaaatcaatttcctcctcgttcattcgtttttatcaatcaatcagtcgctg 69
DB 418 ATGAAATCATATATCTTTTATGCTTCATTTCTTTTATTTAATAACACAAATGTA 477
QY 70 acccaagaatccatcagaagcgtgttaagaactggaagccttgaagaatgcgcctt 129
DB 478 ACACATGAAGAGTTATCAAGAACTGTCAAAAACCTTAGAAGCTTTAGAAAGTCGATTTG 537
QY 130 accggaatcagcgtgttcagaagaagatgctgataatgaagaagcagtcgacg 189
DB 538 ACAGATTATAGTTTATTTCAAAAGGAAAAATGCTTTAATAATGAAGAACAACTGGAACA 597
QY 190 gcgcgttaacaacgaacaccgcgtttctaaaggctcgtgctagcggctgcgcgtg 249
DB 598 GCTGTTCACAACTAGTACACCTGCTTCAAAAGGCTTCAGTTGCTTCAGGTTGCTCAGTG 657
QY 250 tctgtgacctcgtgggtgtccgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 309
DB 658 TCAAGTGTCTTCAGGTGCTTCAGGTGCTTCAGGTGCTTCAGGTGCTTCAGGTGCTTCAG 717
QY 310 gaaagcgcgcgttcgcggagacagtcgaagaacacatccatctgacaacactagcga 369
DB 718 GCTTCAGGTGCTTCAGGTGCTTCAGGTGCTTCAGGTGCTTCAGGTGCTTCAGGTGCTTCAG 777
QY 370 gacgcgaagtcctcgaacgcacactcaagcagagtgagaacatctcctcaatcaag 429
DB 778 GATGCTAAATCTTACGCTGATTTAAACACAGAGTACGAAATTAATCTGTTAACTATCAAA 837

QY 430 gagctgaagtlaccacacagctgttcgacctcctcaatcatatcgtgacctgtgtataac 489
DB 838 GAACATAAATATCTTAACCTTTTGAATTTAACTAATCATATGTTAACTTTGTGTGTAAT 897
QY 490 attcatgcttccaatcatalctgatlgaacggttacgaagaatcaatgaactcctgataag 549
DB 898 ATTCATGCTTTCAAAATATTTAATGATGATGATGATGATGATGATGATGATGATGATGAT 957
QY 550 ttgaattctactcgtacctgtcgaagggcccaacgaaatgaagtggttgccaatgact 609
DB 958 TTAACCTTTTATTTATTTATTTAATTAAGCAAAATTAAGATGATGCTCATGATGAT 1017
QY 610 tgcgaattccatcattcaatttgaagatcgaagcgaagcttgagctatgagaagctg 669
DB 1018 TGTCAATATCCTTTCAATCTTTAAATTCGTGCAATATGAATTTAGACGTACTTTAA 1077
QY 670 gtctcgaatcgcgaagcctcctcgaacacatcaagaacatgtgggaaagaatgagaat 729
DB 1078 GTGTTGGAATATGAAGAACCATTAACATATTAAGATGATGATGATGATGATGATGAT 1137
QY 730 tatataaagaataagaagaccatcgaagaacattacgagctgatacgaagaatccaa 789
DB 1138 TACATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1197
QY 790 aagaccatagacaagaataagaatgcaaccgaaggaagaagaagaagaagttgacag 849
DB 1198 AAACATATTGATTAATAATTAAGATGCACTTAAGAAAGAAAGAAAGAAAGAAAGAAAG 1257
QY 850 gcccaagtlagacctgtcctcctataaacaacagcttgaagaagcccaatcctcagc 909
DB 1258 GCTCATATGATGCTTTCTATTTATTAACATTAACATTAAGAAAGCACTAATTAATTA 1317
QY 910 gtaactgagaagcgcatagacacctcaagaagaatgaaatatacaaaactgtctgac 969
DB 1318 GTTTTGAAGAAACGATTTACACTTTTAAATAAATAAATAAATAAATAAATAAATAAATA 1377
QY 970 aagatataagaatlaagaatcctcgcgcgaagcgaagcttgggaacccctcaacagc 1029
DB 1378 AAGATTAATGAATTAATAAATCCCGCCGCGCAATTTGGAATATACCAATTAATCTGTC 1437
QY 1030 ctggaacaagaacaagaagaatagaagacagcaagaagaatcaagaagaatccgcaaac 1089
DB 1438 CTTGATTAAGAACAAAATAATCGAGGAAACAGAAAGAAAGAAATTAAGAAATTCGCAAA 1497
QY 1090 attaatctcaacatagattcctccttactgatacccttgagagtgagtaacttgaga 1149
DB 1498 ATTAATTTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1557
QY 1150 gagaagaatlaagaatlaagacatcctcgcgaagcgaagcttgaagaagaagaagaac 1209
DB 1558 GAAATAAATAAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1617
QY 1210 aatgaatcccaatgltgacgtlaccctcgtcttataaagatatacaaacgctctc 1269
DB 1618 AATGAATATCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1677
QY 1270 aacgagctcaatagctcgtgtgactgatttaaccccttgatataagaagaacccctc 1329
DB 1678 AATGAATTAATTTCTTTGCTGATTAATTAATTAATTAATTAATTAATTAATTAATTA 1737
QY 1330 aagaatctcagaagaacatlaagaagaagaagttatataagaagaatcaagaagaagc 1389
DB 1738 AAAACATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1797
QY 1390 aaatltgagaagaagaanaatltgagagtgacaagaagaatltagaagaagccgcagaa 1449
DB 1798 AAAATGAAAGAAAAAATTTGAATCTGATTAATAAATCTTACGAAGACGATCTAAGTCT 1857
QY 1450 ctaacgatatataagaagatgtgaaagcgtcgtgaacgagatctatgataccaatc 1509
DB 1858 TTAATATGATTAATAAAGAAATGAAATTAATTAATTAATTAATTAATTAATTAATTA 1917

[illegible]

Db	2998	tgtaattttatataaaccttaagagataatgacaaaattttatattatagacgaacccgaaa	3057
Qy	2650	aagcgtccaccctctgttaaacctctctccgaagtcacgaacacgtctctca	2709
Db	3058	aaagatattccatctctgtaaaaacctctttcaagttcattcattgcacacattatcattnaca	3117
Qy	2710	cctcaagaacgcccggaagtgagcgtcaacgaacacctctcaactcgacccacctat	2769
Db	3118	cctcagatataaacccgaagtaagtcgaataatgatattcattcatttcaatttgaa	3177
Qy	2770	aactcactgaactctgttgagaacatccctctctcgcaagaataagaacatctacaa	2829
Db	3178	aatagttttaaaattattttgaaaaacatttttcagctcttgaaaaaacaaaatttatnccaa	3237
Qy	2830	gaacctatctgacgaataatcgccgagaaactctacgagaataatctgaagaacgagac	2889
Db	3238	gaatttaataaggtcaaaaaaactagtgaaaactttttatgaaaacgtatttttaaaacatagtcat	3297
Qy	2890	acattctataacgagagctctcaactacatctcgtaaacctcaagccgatctgatacaact	2949
Db	3298	acatttttatatgaattctttttacaaattttgttaatttttaaaagctgattgttttaattca	3357
Qy	2950	cttaacgatlgaatctaaacgttaagaagcttgaagaggacatcaataagcttgaagaaca	3009
Db	3358	ttgaattgatgtgattcaaaaagcaaaaattttgaaagaaagatttttaataattaaaaaaaact	3417
Qy	3010	ctgcaacgagctctcgaccctgtcaacaaagtaacaaactgaactgagagactctgac	3069
Db	3418	ttttagctttattatttgattttttatatttaatttaatttaaaatttgcaaaagattatttgat	3477
Qy	3070	aagaagaagacagtcggaacgataaagatgacagatcaagaagctgactctgctcaagag	3129
Db	3478	aaaaagaaaacagttggtaattatataatgcaaatttaaaaaacttactttatttataaagaa	3537
Qy	3130	cagcttgaagcaaacctcaactacgtgaacaatccgaacacgtaactcgagaactctca	3189
Db	3538	caattagaattcaaaaattggaattcactttaaaacccaaacacagttattacaaaaacttttct	3597
Qy	3190	gtgtcttcaacaaagaagaagccgagatctcgcgagacagagaacacactctgagagac	3249
Db	3598	gtttttcttttaacaaaaaaanaagactggaattgacgaanctgtaaaacacattagaaaac	3657
Qy	3250	accaagatctctcacaacatcaaaagccctgcgaattatataatgctgcctctc	3309
Db	3658	acaaaaatatttatgaaaacatttaaaagacctgttaattattatataatgctgaattcatt	3717
Qy	3310	cccttgaagactctcccgagagagcatalccagaccgagataactaagccagctcgag	3369
Db	3718	ccatttaaaaccttttaagtgaaagattcatttcaaacgaaagattatattatccagtttaaaa	3777
Qy	3370	aacttcaaggtccgtcttaagcttcggaaggaacttgaagaacaaactgaacttgaagag	3429
Db	3778	aaatttaaaagatttaagtaatttaagaagaaaaatttaagatataatttttaattttacaaaag	3837
Qy	3430	aagaagctcagctacactctagcgagctgcatcaactgatactcgccgagctcaagaagctc	3489
Db	3838	aaaaaattatcatcttatacgaaggtttacattcatttttaattgtcgaaatttaaaagaafta	3897
Qy	3490	attaaagaacaaactacacccgcaatagcccaagcgagaataataiaacagaogtgaataac	3549
Db	3898	atataaaataataaaattttatacagtaattttcccaagtgaaaaataatacgatttttaacaa	3957
Qy	3550	gcacgtggaattacagaagttctcgctcgaggaagaagatgctgcgaactgtgtgtct	3609
Db	3958	gcatttagatttttcaaaaaaattttctccgaaggaacagattgttgcaacgatttttaagt	4017
Qy	3610	gaatctgctcgcgacacatgagacgcttcaaacctaaagaacgtgcactctcaatgtc	3669
Db	4018	gaatatgattctcgacacattttagaacaaagtcacaaabaaacacagcatcaactatgtta	4077
Qy	3670	gtagccgagttccaatataatccaatctcagaacgtcgacgattgagctgattgagctc	3729

Dh	4078	GGACAGAGCTTAACACATTAACACACTCCACAAATGTCGATGATGACAGTACGTAACGTA	4137
Qy	3730	atcatctgacctatcttcgycgagagcgagagagagcctacagatgacctcggccaggtgctc	3788
Dh	4138	ATCATAGTACCTATATTGGAGATCCGAAGAAGATTATGATGATTAGCACAACTAGTA	4197
Qy	3790	accggtgagcgtctcactctctccgfgatctgataacaatctgtccaaatcgagacgaa	3849
Dh	4198	ACAGAGAGAGCAGTACCTCCCTCCGTAATGTAGATACATCTCTTCATTAATTTGAAAAATGAA	4257
Qy	3850	taagaaatgacctctcgaagacctcgcgagcgctctatagttctctcgaagaagcgctg	3909
Dh	4258	TATGAGGTTTTATATTAAAACTTTAAGCAGGTCGTTTATTAAGAACTTTAAAAAACAAATTA	4317
Qy	3910	gagataacgctgatatccctccaatgtctcaacgctaaagacatctcgaacgcgcctctat	3968
Dh	4318	GAATAATACGTTATGACATTTAATGTTAATAGTAAGATATTTTAATTACAGATTTAAT	4377
Qy	3970	aagagagaaaattccaagaaagctctggagagcgacttgattccctataaagacctgacc	4029
Dh	4378	AAAGGTGAATAATTCAAAAATGTTTTAGAAATCGATTTAATTCATATTAAGATTTAACA	4437
Qy	4030	tcacctactacgtctgcgaagaccatacgaattccctcaataaagagaagagggataaa	4089
Dh	4438	TCAAGTAAATTAATGTTGTCAAAGATCCATATAAATTTCTTAATAAGAAAAAGACATATA	4497
Qy	4090	ttctctgctagttcaactatatacgaagacctcatcgacacccgatacctaactctgctat	4149
Dh	4498	TTCTTAAGCACTTTAATTATTATTAAAGATTCATAGATACGGATATTAATTTTCCAAAT	4557
Qy	4150	gagtgcctggggattatacaagatccctgcagcgaaaaatacaagctcgcacctgactcat	4209
Dh	4558	GATGTTCTTGGAATTTAATAAATAATTATCCGAAAAATTAATAATAGATTTAGATTCATTT	4617
Qy	4210	aaaaagratctcaacgctgaagcagcgagaaatgaaaaaatactgcctccctctgataaac	4268
Dh	4618	AAAAAATATATCAACGCAACAAAGGTAAGAAAAATGAAAAATACCTTCCTTTTAAACAT	4677
Qy	4270	atcgaaaccctgcbaacagacagtgaaacgcaaaaatcgacctctctgatactcaacctgag	4329
Dh	4678	ATTGAGACCTTATATTAACAGTTAATGATTAATGATTAATTTGTTATTCATTTAATA	4737
Qy	4330	gcacaagctcctcaactatactctacgagaagagcaatctggaagcttaaaatcaagaagctg	4389
Dh	4738	GCAAAAGTTCCTTAATTTATACATATGAGAAATCAAAAGTGAAGTTAAAAATAAAGAACTT	4797
Qy	4390	aacctactcaaaaacaaatccgaagcaagcgtcgagcaatcttaagaaaaaataacaattctgc	4449
Dh	4798	AATTACTTAAAAACAAATTCAGACAAATTTGGCAGATTTTAAAAAAAATAACAAATTTGCTT	4857
Qy	4450	gaaatcgagaccctgcttaccgatacttaacacaacaactctctgaccaagctcttgcctc	4509
Dh	4858	GGAAATGCTGATTATTAACACAGATTATTAACCAATTAATCTTATTTGACAAAGTTCCTTAGT	4917
Qy	4510	actggaatggtgtctcgaanaacctcgccaaacagctgctgaagcaatctgctcgaagcgcaac	4569
Dh	4918	ACAGAGTATGTTTTTGCAAAATCTTGCTAAAAACCGTTTTATCTAATTTACTTGATGGAAC	4977
Qy	4570	ctggaagggactgctggaacatctcccgacaccaaatgctgagaagaaacagctgccccagat	4629
Dh	4978	TTGGTAAGGTATGTTTAAACATTTTACCAACCAATGCGTAAAAAAAACATGTCCACAAAT	5037
Qy	4630	agcgccgttctcgaagcactcgcgaagcgcggaagctgcaagctgctcctcctgaaactaaa	4689
Dh	5038	TCCTGAGTGTTCACACACTTATAGATGAAGAAGAAATGTAAATGTTTTATTAATTACAA	5097
Qy	4690	caagaagagataagctgctgcggaacccaacacctaccctgcaatgaaacacaatgycgg	4749
Dh	5098	CAMGAAGGTATTAATGTTGTTGAAAATCCAAATCCTACTTTGTAAGCAAAATATATGTGGA	5157
Qy	4750	tgtgaagccgatacttaatgctacccggaggaagacagcgctctcaacggaaagaaatacaca	4809
Dh	5158	TGTTATCTCAAGATGCCAAATTTACCGAAGAAGATTACAGTACGACGGAAGAAAGAAATTCACA	5217

[illegible]

QY 1990 ccalactacatcgcgtactcaagaagatagacaaactgaaagtgttctacgcccacaa 2049
DB 3572 CCTATTATTATTATTGTTGAAAAAGAAATGATTAATTAAGTTCATGCTTAAG 3631
QY 2050 gtcgagagccgtacacgaagaagaagaacatlaaactcgaaggaagtcagataac 2109
DB 3632 GTAGAAATCATTTGTAATGAAATGAAAAAACAATTAACAGAGTCAATGATTAAT 3691
QY 2110 tccgagccctccacagaagaagataaacgcgacgtacacccagcccgagacacag 2169
DB 3692 TCCGAACCATCAACGAGAGAGAAATPACGACAGCAACTCAAAACCTGGACACAA 3751
QY 2170 gccggtcagctcgcgaagcgatagcgtcgaagctcaagcacaagaagcgagcgagca 2229
DB 3752 GCAGGATGCTGCTTTAGAGAGAGATTCAGTACAGCAACAGCACAAGACAAAAACAGCA 3811
QY 2230 cagctccagctgcagctgcgcgttcagagcgctaaagctcaagtgctcaccacagct 2289
DB 3812 CAACCCAGATACCACTACCAAGTACCAAGCAAAAGCACAAAGTCCCAACACACAGCA 3871
QY 2290 cctgtgaataaagaagccgagaaatgcacgaacgtgactactctgaagagctctatgag 2349
DB 3872 CCAGTAAATTAATAAACTGAAAAATGTTCCAAATTAATTAATTAATTAATTAATTAATTA 3931
QY 2350 ttccgtgaatcacatccatcacatctgcacaaatatatccctgcgtctcagcagacatgaac 2409
DB 3932 TTTTAAATCACTTCAATATATGTCACAAATATATTTGTTTCCACACTCACTATGACAC 3991
QY 2410 gagaagaatcttaagaagtaacgaataacgaagaagagagataactgcctctgt 2469
DB 3992 GAAAAATATTAATAAAATTAATAAAATTAATAAAAGGAAAGAAATTAAGTTATATGT 4051
QY 2470 gatccactgagctcgtgtcgaatcacaatcacatccatccgttctgattctatgctc 2529
DB 4052 GATCCATTAAGCTTATGTTTAAATATACAAATTAACATACCTGTAATGTAATGTTT 4111
QY 2530 gatagcccaacaatctctctctcctcactgctacatgagatatatagaagaagatgctc 2589
DB 4112 GATAGCTTAAACAATGTTTATATCACTCACTATTTATGAAATTTATGAAAAAGAAATGTT 4171
QY 2590 tgcgaactgtataaactcaagaacgaagaagataagaacactctcgtgaagaagctaaag 2649
DB 4172 TGTAAATTTATTAACCTTAAGATATGACAAATTAATAAAATTAATAAGAGGAA 4231
QY 2650 aaggtctccactcgtcttaaaactctctcctcagctcactgcaacacgtctctcaca 2709
DB 4232 AAGATATCCACATCTGTAAAACTCTTCAAGTCAATGCAATGCAACCATTAATTAACA 4291
QY 2710 cctcaagacagcccgagctgagcgctaaagcgaacactctcaactcgaaccaacttaac 2769
DB 4292 CCTCAGGATTAACCCGAAATGATGCAATGATGATACATCTCAATTAATTTGAAT 4351
QY 2770 aactcactgaaactgtttgaagaacatccgtctctcgtgaagaagataagaacatctcaca 2829
DB 4352 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4411
QY 2830 gaactatctgacagaacatcgtccggaacactctcaaggaagatactgaagaagcagcgac 2889
DB 4412 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4471
QY 2890 aactctataacgagagctctcaactaactcgtgaaactcgaagcgatgatacactc 2949
DB 4472 ACATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4531
QY 2950 cttaacgagatcttaagaagtaagcgtgaagaagagatacaataagctgaagaagaca 3009
DB 4532 TTGAATGATGATCAAAAGGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4591
QY 3010 ctgcaactgagctcgaactcgaactgaacaaagtaacactgaagaagactcttcgac 3069
DB 4592 TTACAGTTATCATTTGATTTATATTAATTAATTAATTAATTAATTAATTAATTAATTA 4651

QY 3070 aagaagaagacgctcgcaagtaaaagatgcagatcaagaagtgagctcgtcacaagag 3129
DB 4652 AAAAGAAAAACATTTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4711
QY 3130 cagctgaaagcaactcactcaactcgaacatccgaacacgactcgaagactctca 3189
DB 4712 CAATTAAGAAATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4771
QY 3190 gttgtcttaacagaagaagaagagcgagatcgccgagacaggaacactcgtggaac 3249
DB 4772 GTTTCTTTTACAAAAAAGAAAGCTGAATTAATGCAAAATCTGAAACCATTAAGAAAC 4831
QY 3250 accaagatctctcacaacactcaagaagccctgcataagtatataatgagcagctctc 3309
DB 4832 ACAAAATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4891
QY 3310 cctctgaagaactctcgaagaagagacatccgaaggaataactcgcagcctcgag 3369
DB 4892 CCATTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4951
QY 3370 aacttcaaggtcgtctgaagctcgaagcagctgaagacacactggaactggaag 3429
DB 4952 AACTTTAAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5011
QY 3430 aagaagctcagctcactcctcagcagcagctcagctcagcagcagcagcagcagc 3489
DB 5012 AAAAATTTATCACTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5071
QY 3490 attaagaacagaactcaacccggaatagccgaagaagaataatacagaagctgaataac 3549
DB 5072 ATAAAAATTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5131
QY 3550 gcaactgaaacttcaagaagctcgtcgtgaagaagaagagatgctgcagctcgt 3609
DB 5132 GCATTTGAATTTTACAAAAATTTCTCCAGAGAGACAGATGTTCAACAGTGTGAAGT 5191
QY 3610 gaactggtctcgacacactcgtgagcagctcgaacactcaagacactcactcactc 3669
DB 5192 GAAAGTGATCGACACATTAAGAAAGTCAACCAAGAAACAGCATCACTCACTCACTCA 5251
QY 3670 ggaagccgagctcacaatcaatccacatctcgaagcgtcgaagatgagatgagctc 3729
DB 5252 GGAGCGAGCTTCAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCA 5311
QY 3730 atcatgtgctcactcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 3789
DB 5312 ATCATGTATCACTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5371
QY 3790 accggtgagcgtcactcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 3849
DB 5372 ACAGGAG 5431
QY 3850 taggaagctcctcactcgaacactcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 3909
DB 5432 TATGAGGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5491
QY 3910 gagaataacgtgacactcgaactcgaactcgaactcgaactcgaactcgaactcgaactc 3969
DB 5492 GAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5551
QY 3970 aagaagagaatctcaagaagctcgtggaagcagctgatacctcctataaagaagcagc 4029
DB 5552 AAACGTAAGAAATTTCAAAATGTTTGAAGATCAAGATTTAATTAATTAATTAATTAAT 5611
QY 4030 tcccttaactcgtgtcgaagaagcagcagcagcagcagcagcagcagcagcagcagcagc 4089
DB 5612 TCAAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5671
QY 4090 ttctgctcagtttacaactatacaagaagctcctcgaagcagcagcagcagcagcagcagc 4149
DB 5672 TTCTTAAGCAGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5731
QY 4150 gatgtcgtggttatcacaagatccctgagcggaataacgaagctcgaactcgaactat 4209

Db 901 GTTTTAGAAAAAGTATGACACTTTAAAAAATGAAACATTAAGCAATTAATCTGAT 960
Oy 970 aagattaatgaatataagaatctctccgcgaactctgggaaccccttaacagctg 1029
Db 961 AAGATTAATGAATTAATAATCCCGACCGCAATTCGGAATACACCAATTAATCTG 1020
Oy 1030 ctgcgaagaagaagaatagagagcagagaaagatcaagaagatcgccaac 1089
Db 1021 CTGTGATTAAGACAAAAAATCGAGGACAGCAAAAGAAATTAAGCAATTCGCAAACT 1080
Oy 1090 attaatltaacaatagatctctcttactatgaccccttgagctgagtaacttga 1149
Db 1081 ATTAAATTAATTAATGATGTTATTAATTAATTAATTAATTAATTAATTAATTA 1140
Oy 1150 gagaagaataagaatataagacatcccgcaagtcgagagaagaagatcaacagact 1209
Db 1141 GAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1200
Oy 1210 aatgaatcccaatgltgagcgtacccctctgtctataacgatalcaacaagctctc 1269
Db 1201 AATGATATCCAAATGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1260
Oy 1270 aacgagctcaatagctctgagctgagctgataacccctcgatatacgaagaacccctc 1329
Db 1261 AATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1320
Oy 1330 aagaatctacacagacagatgagaagaagaagttatcaacgaatacgaagaagatc 1389
Db 1321 AAAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1380
Oy 1390 aaatltgagaagaagaatltgagagtgacaagaagaagltcagaagacgagcaaaagt 1449
Db 1381 AAAATGAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1440
Oy 1450 ctaaaagatatacctaagaagatagaagaagcgtctgaagagatctatgatlccaatlc 1509
Db 1441 TTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1500
Oy 1510 aacaataacatgacccgtgcgaactctgcgaagaatgaggaagaagcgtactctcaaa 1569
Db 1501 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1560
Oy 1570 gtgaggaacctgcacacataatcttgcatacctatgagaatcttaagaatactt 1629
Db 1561 GTTGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1620
Oy 1630 gagaagctacccaagctcttaagatataagagagatctctcgtgagaaatctgtctg 1689
Db 1621 GAAAGTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1680
Oy 1690 gagaagaacataagatatacagaagatcctcaagtaagatcgaagaagatcgaagc 1749
Db 1681 GAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1740
Oy 1750 ctgtgtgagaacataagaagaatgagaagaagctgtgttgagaagaagaatctcaaaag 1809
Db 1741 TTAGTTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1800
Oy 1810 gaaataaaacagatgagaagaatcccgaggtctccgaatctgttaagltcaagtgag 1869
Db 1801 GAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1860
Oy 1870 aaggtgtctctcatgacagaatltgagacccaagaagaactcaactctctcgaagac 1929
Db 1861 AAGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1920
Oy 1930 gtgcgagttaaacataataatactgycgaagatgataagcaggaagaataagaagcga 1989
Db 1921 GTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1980
Oy 1990 ccatctactcatcgtactcaagaagaagaatagacaaactgaagctgtctatgcccaa 2049
Db 1981 CCTATTATTATTATGTTGAAAAAAGAAATTAATTAATTAATTAATTAATTAATTAATTA 2040

Oy 2050 gtgcgagccttgatcaacgagaagaagaacattaaactgaagagacgtcagataac 2109
Db 2041 GTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2100
Oy 2110 tccgagccttccacagaagaagataaacggagcaggtctacacccaagcccggaacag 2169
Db 2101 TCGGAACCATTAACCGAGAGAGAAATTAACGAGCAAGCAATTAACAAACCTGAGAACAA 2160
Oy 2170 gccggttcaagctctgagaagcagatagcgtgcaagctcaagcacaagaagcagaagcga 2229
Db 2161 GAGGATTCGCTTTAGAAAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTA 2220
Oy 2230 cagccctcagctgcagctgcgcctccagagagctaaagctcaagctgctcaccacagct 2289
Db 2221 CAACCCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2280
Oy 2290 cctgtgaataaagaagaagcagagatgcaagcaactggaactcttgagaagctctatgag 2349
Db 2281 CCAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2340
Oy 2350 ttcctgataacatctctacatctgcgcacaatatatcctcgtctcacaagcactatgaac 2409
Db 2341 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2400
Oy 2410 gagaagatcttaaaacagtaacagataaacgaagaagaagaagtaaacgtctcctctgt 2469
Db 2401 GAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2460
Oy 2470 gatcaactggaacctgctgtltaacatatacgaagaacatcccgatltatgtatctatgctc 2529
Db 2461 GATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2520
Oy 2530 gataagctcaacaatctctcctcaagctcgaagctgaagatataagaagaagatgagc 2589
Db 2521 GATAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2580
Oy 2590 tgcgaacctgtataaacctcaagaagaagaagaagaagatgaagaacctcttgagaagaagc 2649
Db 2581 TGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2640
Oy 2650 aaggtctccacactctgtltaaaactctctctccagctcgaagcaacactgtctctcaca 2709
Db 2641 AAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2700
Oy 2710 cctcaagaagaagccgaagaatgagcgtcaagcagacactcgaactcgaacactat 2769
Db 2701 CCTCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2760
Oy 2770 aactcactgaacgtgttgagaagatcctgtctcgtcgaagaagaatagaacataccaa 2829
Db 2761 AATAGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2820
Oy 2830 gaactatltgagaagaagaatcgtccgaagactctcagaagaagtactgaagaagcagcag 2889
Db 2821 GAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2880
Oy 2890 acattctaaagagaagctctcaactctcgtgaatctcaaaagccgaatgataactct 2949
Db 2881 ACATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2940
Oy 2950 cctaaagatctcaaaagcgtgaagaagcgtgaagaagcgaatcaaaagctgaagaagaca 3009
Db 2941 TTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3000
Oy 3010 ctgcgaactgagctcgaacctgtacacaagaatcaaaactggaagagactctctgcac 3069
Db 3001 TTACAGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3060
Oy 3070 aagaagaagaagctcggcagatataagatgacatgaagaagtgaactgtctcagaagag 3129
Db 3061 AAAAAGAAAAAGTGTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3120

OY	3130	cagcttgaagcaaacctcaactcactcgaacaatccgaacacgctctcgcgaacttctca	3189
Db	3121	CAATTAGAAATCAAAATTGAATTCCTTAATTAACCCAAACGATGTATTCAAACATTTTCT	3180
OY	3190	gtgtcttcaacaagaagaagagccgagatccgcgagacagagaaacacttggagac	3249
Db	3181	GTTTTCTTTTAACAAAAAAGAGCTGAAATGCGAAGACTGAAACACATTAGAAAC	3240
OY	3250	accaagaattctcttcaaaactcaaaagccctgcgaatataataatgvcgagctctc	3309
Db	3241	ACAAAAATATTATTGAACACTTTAAGAGCTTGTTAAATATTATRAATGCGAATCACT	3300
OY	3310	ccctctgaagactctctcgcgagagacatccgagaccgagataactagccagccctcgag	3359
Db	3301	CCATTAAAACTTTAAAGTAGAAGTCAATTTCAACGAGAGATAAATTAGCCAGTTTAGA	3360
OY	3370	aactcgaagtlcctgltcgaagtcggaagcgaagctggaagaaacctggaactvgagaag	3429
Db	3361	AACTTTAAAGATTAACTAAATTTAGAGAGAAATTTAAAGATAATTTAAATTTAGAAAG	3420
OY	3430	aagaagctcagcttaccctcttagcgagctgcacacctcgtatccgcgcgcgaaggaagtc	3489
Db	3421	AAAAAATATCATCATCTTATATCAAGTTCGATTCATCTTAAATTCGTAATTAAGAACTA	3480
OY	3490	attaagaacaagaactcaccgcggcaatagcccaagcgagaaataatacagacgttgaatac	3549
Db	3481	ATAAAAAATAAAAATTATACAGSTAATTCGCCAGTGAAAAATAATACGATTTAACAT	3540
OY	3550	gcactggaatcttacaagaagtlccctgcctgaaggaacagatgtccgcaatgltgltct	3609
Db	3541	GCAATTAGAAATCTTACAAAAAATTTCCCGAAGAGAACGATGTTGCAACAGTTGTAGT	3600
OY	3610	gaatctggtccgacacacgcgagagcagcttcacacccaagaagccgcgaatctcaatgtc	3659
Db	3601	GAAAGTGGATCCGACACATTAGACAAACAAGTCACCAAGAAACCGCATCACTCATGTA	3660
OY	3670	ggagccgagcttcaatacaatltacacatctcgaacgctgcagatgtagtgcagtagcgtc	3729
Db	3661	GGACACAGAGCTTAACCAATAACACATCATCAAAATGTGATGATGAAGTATGATGACGTA	3720
OY	3730	atactgtgcttatcttcgagcgagagcgagaggaactgaagtaagcctcgccgaagtgctc	3789
Db	3721	ATCTAATACCTTAATTTGGAGATCCGAGAGAGATTAGATGATTTGGCAAGTATGA	3780
OY	3790	accggtgaagcgttcaactccttcgcgtatgtgataacatctctccaatatcgagaacga	3849
Db	3781	ACAGGAAACAGCACTACTCTTCGTAATTTGATTAACATCTTTCTTAATTTGAAATGAA	3840
OY	3850	tacgaagtgtctatactgaaacactctgcgagcgctatagcttctcctaagaacaacgtg	3909
Db	3841	TATGAGGTTTATATTTAAACCTTTAGCAGGCGTTTATAGAAAGTTTAAAAAACATTA	3900
OY	3910	gagataacgtgatgaccttcaatgtcgaagcgtgaagagacatctgaaacgcgccttaat	3969
Db	3901	GAAATTAACCTTAATGACATTTAATGTTAATGTTAAGGATATTTTAAATTCACGATTTAAT	3960
OY	3970	aagagagaatcttcaagaagcgtctgagagcgagctgatatccctataaagacctgagcc	4029
Db	3961	AAACGTAAAAATTTCAAAAATGTTTTAGATACAGATTTAATTCATTAAGAAATTTTACA	4020
OY	4030	tcctctaactagctltgcaagaagccatacaagtlcctcctaataaagagaaagagataaa	4089
Db	4021	TCAAGTAAATATGTGTGCAAAAGATCCATTAATTTCTTAATTAAGAAAAAGAGATTA	4080
OY	4090	tttctgtctagttacaactatatacgaagcctccatccgacacgcgatataatctcgtaat	4149
Db	4081	TTCTTAAGCGATTAATTAATTAATTAAGGATTCAAATAGATACGATTAATTAATTTGCAAT	4140
OY	4150	gagtgtcggggtattacaagatcccgagcggaataatcaagtctgacctgactcat	4209
Db	4141	GATGTTCTTGATATTTATTAATTAATTAATTCGAAAAATTAATTAATTCAGATTTGATTCATTT	4200
OY	4210	aaaaagatatacgaagtaagcgaagcgagaa tgaanaatctgccttccttcgataaac	4269

Db	4201	AAAAAATATATCAACGACGAAACCAAGGTGAAAAATGAGAAATACCTTCCCTTTTAAACAT	4260
QY	4270	atcggaaccctgtcaagaagctgaagcaagaatcgactctgtttaaaccctgg	4328
Db	4261	ATTGAGACCTTAATATTAACAGTTAAAGATTAATGATTTATTTGTATTAATTCATTAGAA	4320
QY	4330	gccaaagctcccaactaacttacttaccgaagaagcaatgtggaagtltaaaatcaaggactg	4388
Db	4321	GCAAAAGTTCTTAATTTATACATATGACAAATCAACGTTGAAGTTAAATTAACAACTT	4380
QY	4390	aactaactcaaaacaatccaaagaacgttgcagatltcaagaaataacaattgcgc	4449
Db	4381	AATTACTTAAAAACAATTCAGCAAAATTCGAGATTTTAAAAAATAATTAACAATTCCTT	4440
QY	4450	ggaattggaagccgctgtaccgatataaccacaacatctctgaccaaagtctgtgcc	4509
Db	4441	GGAATTTGCTGATTTATATCAACAGATTAATACCAATAAATTAATGATTAATGCTTACT	4500
QY	4510	actggcatgtgttcgcgaaaacctgcgaagaacagtcgtgacgaatctgtcgaagcaac	4566
Db	4501	ACAGGTATGCTTTTGTGAAATCTTGCTAAAAACCGTTTATCTTAATTTACTTGATGGAAC	4560
QY	4570	ctcgagggcatgctgtaaacatctcccgacaccaaigtgcgtgaagaagaacagtgccccaagat	4628
Db	4561	TTGCAGAGTATGTTAAACATTTTCACACCAATGCGTAAAAAAACATGTCACAAAT	4620
QY	4630	agcggctgtttccagagcatctggaacagcgcggaagtgcaagtgtctcctgaatacaa	4688
Db	4621	TCTGATTTTTCACACATTTTAAGATGAAGAAAGAAATGTAAATGTTTATTAATTAACAA	4680
QY	4690	caagaagaagaataagtcgtggaagaaccacaacctactctgcaatgaagaacaigtgcgg	4749
Db	4681	CAGAAGGTGATTAATATGTGTTGAAATATCCAAATCCTACTTTTAACGAAATTAATGTGTA	4740
QY	4750	tgtgacgcgcgaigtctaaatgacacggaagaacagcggtcttaacggaagaanaatcaca	4809
Db	4741	TGTGATGCAGATGCGCAATGATGACGAAGAAGATTGAGTAGACACGGAAGAAATTCACA	4800
QY	4810	tgcgagtgactaagccccgactctatccactcttcgaaggatltttgtgccagctct	4868
Db	4801	TGTGAATGTACTTAACCTGATTCCTTAACCACTTTTGATGGATATTTTTCGAGTTCCCT	4860
QY	4870	aatttcctgggcacatctcttcctgcgtgatccatgcatcgtgatccgtgtacagctcatctaa	4929
Db	4861	AACCTTCTTAGAATAATATCTTATTAATACCTCATGTTATATATTAATATACAGTTCAATTA	4920
RESULT	6		
LOCUS	A04562	5760 bp	DNA
DEFINITION	P.falciparum P.195 gene.		PAT
ACCESSION	A04562		13-JUL-1993
VERSION	A04562.1	GI:410754	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
gene			
CDS			

/protein_1a="CA000367.1"
/db_xref="GI:107075"
/translation="MKILIFLCSFLFFIINTOCVHESYOELVLEALPAVLITGVY
LFQEKVNLNMGTSQVATVYTPGSGSVASGSGSVASGSGSVASGSGSVAS
GSGSRNTPNDSSQSDSKVADYLKHVRNLTITKELYPOLDFITNMLTKL
IHGRVYLDIGCEPITNELLYKLPYLDLAKINDCANDYOCOTPNKLIRANEDLYK
KLVGYCPRLDNIKDNGKREDYIKNNKTKITENILDESKTIDKNNKTEEEKK
KLYAOTDLYSTYNOLDEANILSVLEKRTIDULKRENTKELDKINIKNPPRANGS
NTPNLTLDKNNKILEEHRKEITAKTIRNIDSLFDLPILPEYVIREKNMIDISAV
ETKSTENPSPNCPVYPLSYNDINNAALMGSLFGLNPPEYVPEKSNITDNER
KEFIDEMIKYIKTEESPEKSYEDRSKNDITKEKELYNETYDSFNNMIDIL
PEKMGKREYSIKVEKTLHPSTPYSKSHNLKELIKALKMEYDSLNINYEKLEKY
YKNILSIENIETLVNIIKKDEOLPEKRTIKDOKDPYELIPEYDIOYOKVLL
MKKIDELKPOLILKAYELKHNHVNSTROEKDOPYELIYVLEKIDKLPYKAP
SLINEKKNITKIOQSDNSPESTEGITQATQFTPQQAQSLBEDSYOAOAQKQAO
OPAPVPEPAKAOVPPAPVANNKENTSYDLEYKLOFNTSYIOHKTIVSHST
MNEKLKQOYKITEESPEKLSOPCDLNTNIONINPVMSMDSNTIYHNHLYWKE
ETIGVYLLMEIYKEWCVNLKYNKIDNLEAKVRSYVYTLSSMOPSLTSL
POKPEVANDNDHSHSTNNLSIKLEPNILSLGKKNYOLITQKSENNKYLIND
SDPYNNSPTNPKSRAODINSLNDSKRRKLEBDINKLKTQLDLSPLNYKYLAK
RLEKPTKQYKMQIKLTKLLOLESKLSLNSYEDNALONSVFKNKREKALDYL
ENLEKNTILKRVKQYKYNSESSLPDLSEESIQEDNALONSVFNKYLKLEGT
ADONLNEKRSYISRLNHLIAELKEVIRKNNNGNSVANNVDNNAESYKFLG
EGDVATVYVESGSDITLEOQPKPASTVHGAESNTITTSQVDEVDVATLYLPEE
SEEDYDILGOVYGEAVTYSNIDILSIENEVLYTLPAGVARYKLOLENWNT
FNWVKDNLNRFKRRKPNFENVTLSLITKOTSNYVYDPRYKLEKRRKDPSS
YNYIKDSIDTIDINADVLYGYKILSEKKRSDLSIKKINDOGENKTYLPTNLITE
GLIATVYDKIDLEPLHLEAVLWYTERKSNVETKELNYLTIDOTKIDPRKNNNV
TIVATVYDWHNNLTILKSTGVAPNLKLSVYSLNMLKLARYHPTTPKRTMIM
OOSGSKRCDPRDEEPCCKCLNTYOQSSKEVENSNPENCNENGCADACKIEEDSG
AKTICETKCDPCYKPLSTYFCSSNMFGLSPILMLITVSEI"

BASE COUNT	2565 a	630 c	725 g	1840 t
ORIGIN				

Query Match	54.8%;	Score 2706.2;	DB 9;	Length 5760;
Best Local Similarity	72.1%;	Pred. No. 0;		
Matches 3587;	Conservative	0;	Mismatches 1343;	Indels 45; Gaps 3

QY	10	atgaaatcatcttctctcctcgttccattctcgtgttttatacatcaaatactgaatgctg	69
Db	216	ATGAAATCATCTTCTCTCTCTTATGTTCAATTTCTTTTATTAATAATACAAATGTGA	27
QY	70	accacgaatcctatcaagaagcttgaagaacttgaagcttctggaagtgcgcctt	12
Db	276	ACAAATGAAGTATCAAGAACTGTCAAAAAACATAGAACTTAGAAGATGCAATTTG	33
QY	130	accggataacgctctgttccaagaagaagaatgtgtcgtgaatgaaggagacgttgcag	18
Db	336	ACAGGTATATGTTTATTTCAAAAGGAAAAAATGATTAATGAAGAACAAATGTGAACA	39
QY	190	gcgcttacaaccacacacccggtctctaaaggctgtgctgaagctgagctgcgctgg	24
Db	396	GCNCTTACAACTATGATACACCTGGTTCAAAAGGGTTCAAGTTCAGGTGGTTGAGGTGC	45
QY	250	tctgtgacctctggggggttccgtgcctcccggcggcagcgttgcataagtttgctcaatg	30
Db	456	TCAATGTCCTTCAAGTGGCTCAAGTTGCTTCAAGTGGCTCAAGTGGCTCAAGTT	51
QY	310	gcaagcggcgtttccgggagacagtctgaagaacaatccatcttgaacacttagcattcc	36
Db	516	GCMTTCAAGTGGTTCAGCTAATTCAGACAGCTACAAATCTTCAAGTATTCATGATTTCA	57
QY	370	gaagccaaagtcctacgcgaactccaagcaccgagtgagaactatctctcactatcaag	42
Db	576	GATGCTAAATCTTACCTGATTTAAACACAGACAGCAAAATTACTTGTTAATATCAAA	63
QY	430	gaagctgaaggaaccaagctgttgaacctcaactatcatatctgcgaacagctgtgtaac	48
Db	636	GAACCTAAATATCTTCAACTCTTTGATTTAACTATCATATGATTAACCTTTGTGTGATAT	69
QY	490	attcaatgacctcaaatcatctgattgaaggtttagaagaagatcatgaactgaactcgttacaag	54

Db	696	ATTTCATGGCTTTCAAAATATTTTAATTGATGAGATATGAGAAATAATGAATTTATATATAA	755
Qy	550	ttgaaattctaactcgcagctctgaaggccaaactgaatgacgtttcgccaaatgacatc	609
Db	756	TTAAACTTTTATTTTGATTTTATTAGAGCAAAATTAAATGATGTATGTCTTATATGTTAT	815
Qy	610	tgtaaaatccatccaatcttgaagatcagaaccagagtttgagctatctgaagaagttg	669
Db	816	TGTCAAAATACCTTTCAATCTTAAATTCGTGCAAAATGAATTTAGACGTACTTTAAAAACCT	875
Qy	670	gtcttcgagatctgcgaagccctctgcgaacaatcaagagacaatgtyggagaagatgaaat	729
Db	876	GGTTTCGGATTTAGAAAAACCTTAGACAATTTTAAAGATAATGTAGCAAAAAATGGAAGAT	935
Qy	730	tattctaaagaagaaatgaagaccatcgagaaacatttaagagctgtgagaaagccaa	789
Db	936	TACATTTAAAAAAAATTAAGAAAAACCATTGAAAAATATTAATGTAATTTAATTGAAGAAAGTAA	995
Qy	790	aagaccatagacaanaaataagaaatgcacaacaaagagagaaagaagaagttgtaacag	849
Db	996	AAAACAATTTGATTAATAAATTAAGAAATGCAATTAAGAGAAAGAAAAAAAATTTATACCA	1055
Qy	850	gcccaagatcagcgcgtgcactataaacaacagcttgaagaagccatacctcatcagc	909
Db	1056	GCTCATATTTGATTCCTTCTATTTTACAAATTAACAAATTTGAAAGAAACACATATTTAATAAGC	1115
Qy	910	gtacgtggaagcgcatagacaacctcaagaagaatgaaatactcaagaagactgtgcag	969
Db	1116	GTTTTAGAAAAACGTATTGACACTTTTAAAAAAAATGAAAACCTTAGGAATTTACTTGAT	1175
Qy	970	aagatttaa tgaatttaagaatccctccgcagcaacactctggyaacaaccccttaacagctg	1039
Db	1176	AAGATTAATTAAGAAATTTAAAAATCCCCCAGCGGCACATTTTGGAATATACCAAAATCTCTC	1239
Qy	1030	ctcgagcaagaacaaagagatagagagacgcgagaaagagatcaagaagatgccaaaac	1088
Db	1236	CTTGATTAAGAAACAAAAAATCGAGGACACGAAAGAAATTAAGAAATTTGCCAAAATCT	1299
Qy	1090	attaagttcaacatagattctctctcttaccagatcccttgagcttgagatgactacttgga	1149
Db	1296	ATTTAATTTTAATTTGATAGTTTATTTTACTGATCCACTTGAAATTTAATAATATCTATTTAAGA	1355
Qy	1150	gagaagaa taaagaaatagacatctccgcgcaaaagtcgagacaagaatcaacgcgaact	1209
Db	1356	GAAAAAATTAATAATTTTGATATATAGGCAAAAGTTGAAACAAAGAAATCAACTGAACCC	1415
Qy	1210	aatgaatacccaatgltgtgacgtlaccctctgtcttatacagataccaagaagctctc	1269
Db	1416	AATGAAATTCCAAAATGCAAGTTACTTATCTCTGTCAATTAACGATATTAACAAATCTTTA	1475
Qy	1270	aacgagctcaaatagcttcgcgtgactgatttaaaccccttgattatagaagaagaccctc	1329
Db	1476	AATGAACTTAATTTCTTTTGGATTTAATTAATTCACATTTGATTTATACAAAGAACCAAGT	1535
Qy	1330	aagaatctctcacagacaatlgagagaaagatctatacacgaaatcaagaagaagatc	1389
Db	1536	AAAAACATATTAATCTGATTAATGAAAGAAAAAATTCATTAATGAAGAAATTTAGCAAAAAATTT	1595
Qy	1390	aaaattlgagaagaanaaattgagagtlgcagaanaaagttacgaaagccgcagcaaaagt	1449
Db	1596	AAAAATGAAAAAATAAATTTGAATCGATTAATAAATAATTTTACGAAGACGATCTAAGCTT	1655
Qy	1450	ctaaagatctacttaaaggtatgnaaagctgtctgaacagagatctatgatctcaaatc	1509
Db	1656	TTAAATGATTAACAAAAAGATATGAAAAATTTTCTTAAGAAATTTATGATAGCAAAATTC	1715
Qy	1510	aacataacatcgagcccgacaaactcgagaaagatgagggaaaacggtactcttaaaa	1569
Db	1716	AATTAATTAATATAGTTTAACCTAATTTTCGAAAAAATGATGCGTAAAGATATTTCAATATAA	1775
Qy	1570	gtggagaacatgacaaccaataactctgtcatccctagagaaatctlaagcataactct	1629
Db	1776	GTTGGAACACTTACAGCCCTATAGCTTTTGATTTGATCTCTATGAATAATTTTAACATATATCTT	1835

OY	1630	gagaagctc	ceccaaagctctctaa	gatabgagactatctctc	gcggaacatgctg	1689					
OY	1630	gagaagctc	ceccaaagctctctaa	gatabgagactatctctc	gcggaacatgctg	1689					
Db	1836	GAAAGATTAAACAAAAGCTCTTAATATATGCAAGATATTCTTTAAAGCAATATAGTACTT	1895								
OY	1690	gagaaagacacaaagattctaa	gaatctccataagtaaga	tcgaaacggaagtcgagcg	1749						
Db	1896	GAAAAAGCAATTTAAATATTTTAAAAATTTTATATAGCAAAATAGAAATAGCATTTAAACA	1955								
OY	1750	ctctgtga	aacatlaaagaaga	tgaagaacgctgtgtgtga	agaagattcaaaagac	1809					
Db	1956	TTAGTTGAAAATATTATAAAAAAGATGAAGAACACCTTTTGAAAAAAAATTTACTAAAGC	2015								
OY	1810	gaataa	ccaagctggaagatcccg	ggagctccgatatgtta	agtcgaagtcgag	1868					
Db	2016	GAAATATTAACCCAGTGTGAAAAAATTTTATGAAAGTATCTGCAATTTGTAAAGTACAACTTTCAA	2075								
OY	1870	aagg	tgctccctc	atgaacaagatttga	actccaagaagacccaactatctc	tgaagac	1929				
Db	2076	AAAGTTTTATTAATGACCAAAATTTGACGAATTTAAAAAACCTCATTGTATTTAAAAAT	2135								
OY	1930	gtgga	gttaaaacataata	ctacatgctgcg	aatagltataagcag	ggagataagcagga	1989				
Db	2136	GTAAGATTTAAACCTATATATACATGTCTCCAAATTTTCAAAAAGAAATATAGCAAGAA	2195								
OY	1990	ccat	ctacacccatc	gcgaacccaagaagata	gaacaacgtaaa	gcttctgcgcaa	2049				
Db	2196	CTTATTTATTTAATGTGTGAAAAAAGAAATTTGATTAATTTAAAAAGTGTTCATGCTTAAAG	2255								
OY	2050	gtcga	gagcctg	atcaacgaagaaga	agaacatlaa	acttgaagcagtcagataac	2109				
Db	2256	GTAAGATTCATGTTAAATGTAGAAAAAATAAATCATTTAAACCAAGCTCAATCGATAT	2315								
OY	2110	tcga	agcctcccaagaaga	gagataaacgcga	gcagctcacaacga	gcccggaacaag	2169				
Db	2316	TCGGAACCATTCACCGAAGAGAAATATACAGACACAACTTACAAAACCTGGACACAA	2375								
OY	2170	gcgcg	gttcagctc	tcgaagcgatgcgtg	caagctcaagcaagaag	cagaagcagga	2229				
Db	2376	GCAGAGATCTGCTTTAGAGAGATTCAGTATCAAGCCACAAAGCAAGAAACAAAACAAACA	2435								
OY	2230	cagc	ctcgaagcgccg	gtctcgaaggtc	aaagtc	caagtgctc	ataccacacagct	2289			
Db	2436	CACACCACAGTACAGTACCAAGTACCAAGAACCAAAAGCCAAAGTCCACACACACACACA	2495								
OY	2290	ccgt	tgataa	caagaccggaatg	tcacga	caactgacactg	gagaagctctatag	2349			
Db	2496	CCAGTAATATATAAACTGAAATGTTCCTCAATTTGATTTCTTGAAAAATTTATATCAA	2555								
OY	2350	ttcct	gata	ctacatctac	ctcgc	caaatatata	ctctgcctc	atacagactaagac	2409		
Db	2556	TTTTTAATACCTTCATATATATATATCTCAAAATATATTTTGTTTCCACCTCAACATATGAC	2615								
OY	2410	gaga	gaatctct	aaacagatacaaga	ataaccagaaga	ggagagaa	gtaaa	ctgctcgt	2469		
Db	2616	GAAAGATTTAAACCAATATATAAATTTTACAAAGAGGAAGAAAGCAAAATTAAGTTATCATGT	2675								
OY	2470	gata	ccactg	gac	ctgctca	atctccga	acaatctcc	ggtata	gctctatgct	2529	
Db	2676	GATCCATTTAGACTTATGTTTATATATACAAATATACATCTGTAATGTATTTATGTTT	2735								
OY	2530	gata	gctc	taaca	-----	atctc	2547				
Db	2736	GATAGCTTTAAACATAGTTTATATCAACAATTTATGGGTTTATGAAAAAAATAATGGTTAT	2795								
OY	2548	ctc	ctc	taactg	ctc	atga	gata	atata	gataaactc	2607	
Db	2796	GTAATTTATTAATGAGCAAAATTTATGAAAAAAGAAATGTTTGAATTTATATAAACTT	2855								
OY	2608	aaaga	caaga	caagataa	gaac	ctctc	ggaggaag	taagaag	gtctc	caactgct	2667
Db	2856	AAGATATATGACAAATATATAAATTTTATTAAGAGGAAGCAAAAAAGTATTCACATCTGTA	2915								

QY	2668	aaactctcttccagctcccaagcaacacgtctctctcaacacctaaagacccgaa	2722
Db	2916	AAAACTCTTTCAGTTCATCATATGCAACCACTTATTCATTAAACCTCAGATTAACCCGAA	2975
QY	2728	gtgagcgttaacagcagcaactcactcgtgccaaccttaataactcaactgaactgtt	2787
Db	2976	GTAACTGCAAAATGATGATACATCCATTCACAAATTTGAATATAGTTTAAATTAATTT	3035
QY	2788	gagaaacccctgtctccgcgaagaataaagaacatctaccagaactctctgacagaa	2847
Db	3036	GAAAACATATTGACTCTTGGAAAAAACAAAATATATTACCAAGAAATTATAGCTCAAAA	3095
QY	2848	tcgtccggaactctcaagaaagatacttgaagaagcagcaacatctatacagagac	2907
Db	3096	AGTAGTGAAACCTTTATGAAAAGATATTTAAAGATAGGATACATTTATTAATGAATCT	3155
QY	2908	ttcactactctgfgaaatctlaaagccgagatataactctcttcaagatgaactaa	2966
Db	3156	TTTACAAATTTTGTAAAAATCTAAAGCTGATGATTAATTAATTCATTGAATGATGAATCAAA	3215
QY	2968	ctgaagagcctggaagagagacatcaatgaactgaagaagacactgaaactgagctgcac	3022
Db	3216	ACGAAGAATATTAGAAAGAGATATTATTAATTTAAAAAAACCTTTACAGTTATCACTTTGAT	3275
QY	3028	ctgtacaacaagtacaacttgaagactctccgacaagaagaagacagctgcgc	3087
Db	3276	TTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3335
QY	3088	aagataagatgcagatccaagaagtgtgacctgtctcaagagagagcttgaagcaactc	3147
Db	3336	AAATATTAATAAGCAAAATTTAAAAAATTACTTTTAAAAAGAACAAATTAATCAAAATTTG	3395
QY	3148	aactcagyaacaa tccgaacaacgtactctgcaagactctccagtgctctccacaagaag	3207
Db	3396	AATTCACTTATTAACCCAAAGCATGATTTCAAAAACCTTTTGCTTTCTTTTACAAAAAAA	3455
QY	3208	aaggaagccgagatcgcgcgagacagagacactctggaagacacaagaattctctcaa	3267
Db	3456	AAAGAGCTGAAATAGCGAAGAACTGAAAACACATTTAGAAAACAAAATTAATTAATTTAGAA	3515
QY	3268	cactacaagaagccctgcgaagtatttaagggagatcttccctctgaagactctccc	3327
Db	3516	CATTATTAAGGACTTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3575
QY	3328	gaggaagcactccagaccggagataactaagccagagctcgagaactcaagctcgtct	3387
Db	3576	GAGAGATCAATTTCAAACAGAGATTAATTAGCCAGTTTGAAAAACTTTTAAGATTAATTAAGT	3635
QY	3388	aagctcgaagcgaagcctgaagaagacaacttgaagcttgaagaagaagagctcagctac	3447
Db	3636	AAATTAGAGGAAATTAATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3695
QY	3448	tctagcagactgcataccctgcgtccgcggagctcaaggaagtcaatcaagaagaagctac	3507
Db	3696	TCAAGAGCTTTTCAATCAATTTAATTTGCTGAATTTAAAGAAATTAATTAATTAATTAATTAAT	3755
QY	3508	accggaactagcccaagcggagaaataaagaacgtgaataagcaacttgaatctttaaag	3567
Db	3756	ACAGGTAAATTTCTCCAAACGCTTAATTAATACGAGTGTTAACAATATGCTATTAGAACTTTACAA	3815
QY	3568	aagctcgtccctgaagaagacagatctgcgcgaactgtgtctctgaactctgctccgacaca	3627
Db	3816	AAATTTTGTCCCAAGAGGACACAGATGTTTGCACAGTTGTAAAGTGAAGAGTGAGATCCGACACA	3875
QY	3628	ctggaagcagttcaactaaagaagacctgtcatctactatctgtcggagccgagtcacaataca	3687
Db	3876	TTAGACAAATCTCAACCAAGAAACCAAGCATCAATCAATGATGAGGACAGAGTCTTAACACA	3935
QY	3688	attaccacatctcgaagcctcgaagcttgagtgctgaatgaatcatatgttgctactctc	3747
Db	3936	ATTAACAATATACAAAATGTGCGATGATGAAGTATGATGACGATATCAATGATCTACTATATTT	3995
QY	3748	ggcagagacgcgagagacatcgatgcacctgcgcgaagtgatcaacggtgaagcttcaact	3807

OY	250	tctgtgscctctgggggttccgtgcgctccgcgcgcgaagcgtggcataagctggctcagtg	309
Db	456	TCAGTTCGCTTCAGGTGGCTCAGTTGGCTTTCAGGTGGCTGCAGTTGCCTCAGTT	515
OY	310	gcaagcgccggttctccggaaacagtcgaagaccacatccatccgaacactcagcattcc	369
Db	516	GCTTCAGGTGGTTTCAGGTAATTTCAGACGTCACAAATCCTTCAGATATATTCAGATATTC	575
OY	370	gaacgcgaagctccatcgcgcgacctcaagaccgagtlgagaaatcatctcctcaatcaag	429
Db	576	GATGCTAAATCTTTCACGCTGATTTAAAAACAGAGTACGAATATACCTGTTTACATCAAA	635
OY	430	gaagctgaagtaecccaagctgtctgcacctcaatcatatcagtctgaacactgtgataac	489
Db	636	GAACCAAAATATCCTCAACCTTGTGATTTAACTAAATCAATGTTAACTTGTGTGATTAAT	695
OY	490	atccatgcttcaaatatctcgatctgcgcgttaagaaagataatgaactccctgcagag	549
Db	696	ATTCATGGTTTCNAATTTTAAATTTGATGATATGAAGAAATTAATGAATTTATATATAA	755
OY	550	ttgaattctctacctcgactctgcttaagggccaaactgaatgaactgttgcacatgactat	609
Db	756	TTAAACTTTTATTTTGATTTATTAAGACGAAATTAATGATGATGATGCTATATGATTAAT	815
OY	610	tgctcaatctcatcgaatttggaagatcagaagccacagctlygacglatlgaagaagctg	669
Db	816	TGTCAAAATACCTTTTCATCTTAAATATTCGCGCAAAATGAATTTAGAGTACTTTAAAAACTT	875
OY	670	gtcttcgatactgcgaagcctctccgcaaatcagaatgaagaatgctgaggaagtlgaagat	729
Db	876	GTGTTCCGATTTAGAAAAACCATTAGACATATTTAAAGATATATGTAGGAAAAATGGAAAT	935
OY	730	tataataaagaataaagaagaccatcgcgaacattaaagagctgatacgaagaatcccaaa	789
Db	936	TACATTTAAAAAATAAAAAACCATAGAAATATTAATGAATTAATTTGAAGAAAGTAAAG	995
OY	790	aagaccatagacaataataagaatgcacaaccaaggaagaaagaagaagtgltacag	849
Db	996	AAAACAATTTGATTAATAAATTAAGAAATGCACTAAAGAAAGAAAGAAAAAATTAATACCA	1055
OY	850	gcccgagcagccgcgttcacatctaaacaacagctggaagaagccatacctcaacgcg	909
Db	1056	GCTCAATATGATCTTCTTATTTTCAATTAACATTTGAAGAAAGCACTATATTTAATTAAGC	1115
OY	910	gtactggaagaagcgcatagacacccctcaagaagaaatgaaatcatcaagaagactgtcgac	969
Db	1116	GTTTTAGAAAAACGTATTTGACACCTTTAAAAAAAATGAAACATTATAGSAATTTACTTGAT	1175
OY	970	aagattaaagaaaltaagaaatcctctcgccagcccaactctggagacacccctcaacgcgtg	1029
Db	1176	AAGATTAATGAATTAATAAATATCCGCCACCGCCCAATTCGAAATATACCAAAATACTCTC	1235
OY	1030	ctgacaagaagaagaagatagaagggccgcgcgaagaagataatcaagagatcgcgaacac	1089
Db	1236	CTTGATTAAGAACAAAAAATCGAGGAACGCAAAAAAGAAATTAAGAAATTTGCCCCAAACT	1295
OY	1090	attaagtccaacatagatctctctcttactgatacccttgagctgagatctactctgga	1149
Db	1296	ATTTAAATTTAATATGATPAATTTATTTACTGATCCACTTGATTTAGATATCATATTTAAGA	1355
OY	1150	gagaagaataaagaatatagascactccgcgaagctcgagacacaagaagatcaacgcgaact	1209
Db	1356	GAATAAAATATAAATATTTGATTAATAGTCGAAGAGTGAAGAAAGAAAGCAATCAACTGACCC	1415
OY	1210	aatgaatatcccaatgtgtgacgatacctctgtcttataaagatatcaacaagaagctctc	1269
Db	1416	AATGAATATCCAAATGGAGTTACTTATCTCTGTCTCATTAATACGATATTAACCAATCTTTTA	1475
OY	1270	aagagctcaatagctcggtagactgataaacccttgatattatcagaaagaacacctt	1329
Db	1476	AATGACTTAATTTCTTTTGGTGATTTAATTAATCAACTTTGATTTATTAACAAAGACCAAGT	1535
OY	1330	aagaatatctcacacagacaatlgagagaagaagttatcaacgaataccaaggaagaatc	1389

Db	1536	AAAAACATATCTACTGTAATATGAAAGAAAAAAATTTCTATAAGAAATTTAGGAAAAAAATTT	1595
Qy	1390	aaatcttggaaagaagaatcttgagatgacaaagaagctacgaagccgacgaagct	1449
Db	1556	AAAATAGAAAAAATAAATTTGAAATCTGATAAAAAAATCTTACGAGACAGACTTAAGCT	1655
Qy	1450	ctaaacgcatctactaaagaatctgaagaagctctgaacgagatctatgatctcaattc	1509
Db	1656	TTAAATGATATTAACAAAAGAAATATGAAAAATTTACTTAATGAATTTTATGATAGCAAAATTC	1715
Qy	1510	aacaaatacatcgaccctggaccaacttcgagaaaaatgattggaaaaacgylactcttaca	1569
Db	1716	AATAATATATATGATTTTAATCTTAATTTTCGAAAAAATATGGTAAAAAGATATTTCAATATAA	1775
Qy	1570	gtggagaacatgcagacacccaataactcttgcatacctgaagaattctcaagaatactt	1629
Db	1776	GTGGAAGAACTTACACACCTTAATTAATTAATTTTGCATCTGTAAGAAAAATTTCTAACATATCTT	1835
Qy	1630	gagaagctcaccaaagctctcttaagatatalgagagacattctctcggaaacatttgctg	1689
Db	1836	GAAAGGTTAAACAAAAGCTCTTAATATATATGAAAGATTAATTTCTTAAAGAAATTAAGTACT	1895
Qy	1650	gagaaagaactaaagatattctcaagaatctcataagttaagatcogaacacgagatcgaga	1749
Db	1866	GAAAAAGATTTAAAAATTTTATTAATAATTTTATTAAGCAAAATGAAAAATGAGATTTAAACA	1955
Qy	1750	cttgcttgagaacatctaaagaagatctgaagaacagctgtttggaagaagaatctcaagaag	1809
Db	1956	TTAGTTGAAATAATTTAAAAAAGATGANAAGACGCTTTTGGAAAAAATAATTACTAAAGAC	2015
Qy	1810	gaaaaataaccagaatgaagaatccctggaagctcccgatatgttataagctccaagctga	1869
Db	2016	GAAATTAACACAGATGAAAAAATTTTATGAAGATATCTGACATTTGTAAGATTAAGTAACTCA	2075
Qy	1870	aagtgctctctcatatgaacaaatgtatgaactcaagaagactccaactcatcttgaagac	1929
Db	2076	AAAGTTTTATTAATGAACAAATTTGACGAATTTAAAAAACATCAATTGATTTTAAAAAAT	2135
Qy	1930	gtggagtttaaaacataatacatcttgcgcgaatagtctaaacagagagaataagaagaa	1989
Db	2136	GTAGATTTAAACCTAATATATATCATGTTCCCAATTTCTTACAAACAAGAAATATAGCAAGAA	2195
Qy	1990	ccatactaccctcatcgatctcaagaagaagatatagaacaaactgaagcttctacgccaa	2049
Db	2196	CCTTATTAATTTATTTGTTGTAAGAAAAAAGAAATTTGATTAATTTAAAGTGTTCATGCTTAAG	2255
Qy	2050	gtcgaagccttgatcaacgaagaagaagaacattaaactgaagagacgtcagataac	2109
Db	2256	GTAGATATCATTTGATTAATTTGAAAGAAAAAATAATTAACCAAGCAAGTCAATCGGATTAAT	2315
Qy	2110	tcgagagcttccaaagaagaagagataaaccggcagagctaccacaaagcccgagaacag	2169
Db	2316	TGGAACCATATACCGAAGGAATTAATTAACGCAAGCAACATTAACAAATCTGAGAACAA	2375
Qy	2170	gcgcgcttcagctctcgaaggcgaatgcgttcgaagctccaagcacaagacagagacagga	2229
Db	2376	GCAGATCTGCTTTTGAAGAGATTTCAGTATCAAGCAACAGCAACGAACAAAGAACAAACAA	2435
Qy	2230	cagcctccagtgccagtgccggttccagagagctaaagctcaagtgctctacaccacagct	2289
Db	2436	CAACACACAGTACAGTACCAAGTACCAAGCAAGCAAGCAAGCAAGTCCCAACACACACACAA	2495
Qy	2290	ccttgagataacaaagaccgggaatgttcagaaactctgactactactcttgagaagctctatg	2349
Db	2496	CCAGTAAATTAATTAACCTGAAAAATGTTTCCAAATTAATTAATTTTGAAGAAATTAATATA	2555
Qy	2350	tctcgaatacatccatacatctgcacaaatataatcctgcgtcttccaaagcactatgaac	2409
Db	2556	TTTTTAAATTAATCTCATATATATGTGCAAAATATATTTTGCTTCACTCAACATATAGAC	2615
Qy	2410	gagaagatctctaaacagatacaagataaccaagaagaagagagataaactgtcctctgt	2469

Db 2616 GAAAGATATTTAAACAAATATAAAATTACAAAGGAGAAAGAAACAAATTAAGTTCATGT 2675
QY 2470 gatccactgacctgctgttcaatcagaaacattccgttatgtatctatgctc 2529
Db 2676 GATCCATTGACTATTGTTTAAATATACAAAAATACATACCTGTAATGTATGTAGTTT 2735
QY 2530 gatagcctcaaca-----attct 2547
Db 2736 GATAGCTTAAACATAGTTTATATCACAACATATTTATGGTTTATGAAAAAGAAATGGTTAT 2795
QY 2548 cctctcactcgttctatgtagataatgagaagagatgctgcgaacctgataaactc 2607
Db 2796 GATTTATATTTACTTATGGAATTTTGAAGAAATGCTGTGTTATATTAATTAACCT 2855
QY 2608 aagacaagaagaagatlaagaacctctcggaagactgaagaagctccacctctgtt 2667
Db 2856 AAGGATTAATGACAAATTTAAATTTATAGAGAGACGAAAAAATATCCATCTGTGA 2915
QY 2668 aaaactctctctccagctccatcgaacacactgctctcacaacctgaagaacacccgaa 2727
Db 2916 AAAACTCTTTCAAGTTCATCAATGCAACCATTTATCATTTAACCTCAGATTAACCCGAA 2975
QY 2728 gtagagcgtcaagaagacacctctcaactcgacaaccttaatactcaactgaactgtt 2787
Db 2976 GTAAGTGCAAAATGATATCATCATCTACAAATTTGAAATTAATGTTTAAATTAATTT 3035
QY 2788 gagaacatctctctctcggaagaataagaacatctacagaagactttatgacagaa 2847
Db 3036 GAAACATATTTGAGCTTGGAAAAACAAATATATACCAGAAATTAATAGTCAAAA 3095
QY 2848 tccgtccgaagactctcaagagaatactgaagaagacagacacatctcatalaagagagc 2907
Db 3096 AGTAGGAAACCTTTATGAAAAAGATTTAAAGATGATGATCATTTTATATGAATCT 3155
QY 2908 ttcaactaactcgtgtaaatctaaagccgagtatatacactctcttaacagatgaactaa 2967
Db 3156 TTTACAAATTTTGTAAATCTAAAGCTGATGATTAATTAATTCATTTGATGATGATCAAT 3215
QY 2968 cgtlaagaagctggaagagacatcaatgaagcgaagacactggaactcgagactcgagc 3027
Db 3216 AGGAACAATTTGAGAACAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3275
QY 3028 ctgtacaacaaglacaaactggaactcggaagactctcgacaagaagaagacagtcgagc 3087
Db 3276 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3335
QY 3088 aagtatagaatcagaatcagaagtgtagtctcgaagagagagcttgaagaacac 3147
Db 3336 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3395
QY 3148 aactcaactgaacatcgaacacagctacgcgaagactctcagtgcttctcaacaagaag 3207
Db 3396 AATTCACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3455
QY 3208 aaggaagccgagatcgcgagagagaagaactctggaagaacacgaactcctcctcaaa 3267
Db 3456 AAGGAAGCTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3515
QY 3268 caactcaagaagcctcgtcgaatataatgagagctctcctcctcgaagactcctcctc 3327
Db 3516 CATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3575
QY 3328 gagagagacatcagagacagagataactcagcagcctcgagaactcgaagctcctcgtct 3387
Db 3576 GAGGATTCATTCACAAACAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3635
QY 3388 aagctcgaagcgaagctgaagacacactggaactcggaagagagaagagactcgaactc 3447
Db 3636 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3695
QY 3448 tctagcgcagctcatcactcgtatcgccgagctcaaggaagatcaatlaagaacaagaactc 3507
Db 3696 TCAAGGAGTTTACATCATTTAATTTGCTGAATTAATTAATTAATTAATTAATTAATTAAT 3755

QY 3508 accgcgaatagcccaagcgagagataatcagaacgtgaataacgactggaactcttaacg 3567
Db 3756 ACAGTAATTTCTCAAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3815
QY 3568 aagtlcctcgtcgaagaagacagatgctgcacactggtgtctcgaatcgtgctcgcagaca 3627
Db 3816 AATTTCTCCCAAGGAAGACAGATGTTGCACAGTTGTAAGGAAGGTGATCCGACACA 3875
QY 3628 ctggagcagtlcacaactaagaacctgcatactcaatgctcggagcgagtcacaataca 3687
Db 3876 TTAGAACAAGTCACCAAGAAACCCAGCATCAACATGATGAGACACAGTCATCAACA 3935
QY 3688 attacacacatcacaagctgacagatgaggtcgaatgagatcaatctgtcctatctc 3747
Db 3936 ATBACACATTCACAAATTTCCATGATGAAGTAGATGACGTATATCTATCTATATTT 3995
QY 3748 ggcgagagcgagagactacatgacctcgccaggtgtcaccggtgaagctgtcaact 3807
Db 3996 GGGAATCCGAGAAAGATTAATGATGATTAAGGACAAAGTAGTAACAGGACAGCAGTAAT 4055
QY 3808 ccttcgtgatgataactctgtccaaaatcgagaacgaatacgaagtgctctatctg 3867
Db 4056 ACTTCGTAAATGATTAACATTTCTTCTAAATTTGAATGAATGATGAGGCTTTATATTTA 4115
QY 3868 aaacctcgtgcagcgctcatalagctctcgaagaacagctggaagataagctgtgacc 3927
Db 4116 AAACCTTTAGCAGGTGTTTATGAACTTTAAACAAATTTGAAATTAATACCTTATGACA 4175
QY 3928 ttaaatgtcaacgtggaagacatctgaaacagccgtttaataagaagaaaaattcacaag 3987
Db 4176 TTTAATGTTAATGTTAAGGATTAATTTAATTAATTAATTAATTAATTAATTAATTAAT 4235
QY 3988 aagctcttggaagagacttgatccctataagacactggaactcctcctaactgcttc 4047
Db 4236 AATGTTTGAATTCAGATTTAATTCATTAATTAATTAATTAATTAATTAATTAATTAAT 4295
QY 4048 aaggaaccatacaagtlcctcaataagaagagaggaataaattcgtcctaagatacaag 4107
Db 4296 AAAGATCCATTAATTAATTTCTTAATTAAGAAAAAAGAGATTAATTTCTTAAGCAGTTATAT 4355
QY 4108 tatataagagactcgaacacagatalcaaatcttcgtataatgagtgtcgtggaattac 4167
Db 4356 TATATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4415
QY 4168 aagatcctcgaagaaaatacaagctcgacctgactcctataaagaatatacgaat 4227
Db 4416 AATATTAATTCGGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4475
QY 4228 aagcaagcgagaatgaaataatctgcctcctcgtgaataacatcgaaacctgtlacaag 4287
Db 4476 AAACAAGGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4335
QY 4288 acagtgaacgacaaaatcgacctcttgtaattcaactcgtggaagcgaagtlcctcaactat 4347
Db 4536 ACAGTTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4395
QY 4348 acttcgagaagaagcaatgtagaattaaatcaagagcgtgaactactcctaacaatc 4407
Db 4596 ACATTAATGAATAATCAACCTAGAAAGTTAAATTAATTAATTAATTAATTAATTAATTAAT 4465
QY 4408 caagaacagctgagatctcagaagaaaatacaaatctcgtcgaatcgcaagctgtct 4467
Db 4656 CAAGCAAAATTTGGCAGATTTTAAAAAATAATTAATTTCTGTAATTCGATATTAATTAAT 4715
QY 4468 accgatltataccacaacaactcctcgtacaaagtlcgtccactcgtgacatggtgtlctgaa 4527
Db 4716 ACAGATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4775
QY 4528 aacctcgcacaaaagctcgtgacatcgtcgtcga-cggcaacctcgcagcgagcgtcgtgaa 4586
Db 4776 AACTTGCTTAATTAATTCGTTTATCTAATTTACTTTGATTTGGAACCTTGCAAGGATTTAA 4835

DB	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	FEATURES
OY	4587	ccctccccaagaccatctcgtgaagaacaacagtg--ccccagaataagcgctgttccag	PFPI90	5282 bp	DNA	INV	21-FEB-1996											
DB	4836	caftttcacaaaccacatctcgtaaaaaaaacaaatgacacagagttcttgagatgttttcaga																
OY	4645	catctggacgacgcggaagatgtaacagtgctccctggaactcaacaagaagaagataag																
DB	4896	caftttagatgaaagaagaagaatgtaaatgtttatttatttaattacaaacaaagagtgatgaa																
OY	4705	tgcgtggagaacccaacccctacctctgcaatgaaacaatgscgggtgtgacgcgcatgct																
DB	4956	tctgttggaaaatttaataatcctactctgttaacgaaaataatggtggatgtgatacgaattcc																
OY	4765	aaatgcaccgcggaagaaacagcgcgtcttaacggaagaataacatcgtcagtgtaactag																
DB	5016	aaattttaccgaagaagaatgtaacgttgcacaaacgaaagaaaacacacatgtgaatgttactaaa																
OY	4825	cccgactcctatcacactcttcgcagcggaatttttgcctcacgctcaattctctggaatc																
DB	5076	cctgattgtgtatccactttccatggtgtaatttttctcagcttcccttcttctttagaata																
OY	4885	tccctctcgtgatcctcaltgctgtaaccctgtaacgattcataataagatgcatg																
DB	5136	tcattcttatttattatcctcatgttatattatattatncaactttcatttaaaaatgtagc																
RESULT	8																	
LOCUS	PFPI90	5282 bp	DNA			INV	21-FEB-1996											
DEFINITION		P.falciparum gp190 (MSAL, MSP1, PMMSA) gene for merozoite surface antigens precursor.																
ACCESSION		X03371																
VERSION		X03371.1																
KEYWORDS		gp190 gene; merozoite surface antigens; MSAL gene; MSP1 gene; PMMSA gene.																
SOURCE		malaria parasite P. falciparum.																
ORGANISM		Plasmodium falciparum																
REFERENCE		Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.																
AUTHORS		1 (bases 1 to 5282)																
		Mackay,M., Goman,M., Bone,N., Hyde,J.E., Scaife,J., Certa,U.,																
		Stuenkelberg,H. and Bujard,H.																
		Polymorphism of the precursor for the major surface antigens of Plasmodium falciparum merozoites: studies at the genetic level																
		EMBO J. 4 (1985), 3823-3829 (1985)																
		2 (bases 1 to 5282)																
		Pan,W., Tolle,R. and Bujard,H.																
		A direct and rapid sequencing strategy for the Plasmodium falciparum antigen gene gp190/MSAL																
		Mol. Biochem. Parasitol. 73 (1-2), 241-244 (1995)																
		3 (bases 1 to 5282)</																

Query Match	Best Local Similarity	Matches 3541;	Conservative	0;	Mismatches 1362;	Indels	27;	Gaps	1;
10	atgaaatcatttctcctcgtcatttctgttttattcatcatcactcaagtgcgtg	69							
72	atgagatcatatcttcttattgctcatttcttatttataaataacaaatgtcta	131							
70	accacgcgaatctatcagagagctgtttaagaactgtgaagctttggaaagatgccgctt	129							
132	acacatggaagattatcaagaacacttctcaaaaaacatgaaagcgttaagaaagatgattatg	191							
130	accggaatacagcgtttccagaagaagaagaatgtgtgcgaattgaagaaggaaagagtgagcag	189							
132	acaggtttatgtttatttttcatatagaagaaaaaattgatctttaattgaagaagaaattt	245							
190	gccgttaacaccagacacaccggttctctaagggtctgtgtcgaatgcgtgtcccggtg	249							
246	-----acrtcaaaaggtgacgaagtgctcaaaagtgatgaagaagtgct	284							
250	tctgtgacctctgggggttccgtgcgctccggtgcgagcggtgtgcaatcaagtgtgcacgtg	309							
285	acaaatgtgataaagtggtcacaagtggtcacaagtgatgaacatgcatatcttgcgtttcaaac	344							
310	gcaagcggcgtttccgggaaacagtcgaagaacaaatcatcttgaaacactatagcattcc	369							
345	actttacctcgttcaaatattcttcatctgtgtgacaaagccctccagctgattgacaaagcattca	404							
370	gaagcgaagtctacgcgcgaacctcaagcagcagtgtagaagaactatctcctcactatacaag	429							
405	gatcttaaatattcgtacgtgattttaaacaacagagacgaatttactgttaaacattatcaaa	464							
430	gagctgaagtaccacagatgttgcgaactcaatataatagtgcgaacggtgtgtataac	489							
465	gaactcaaatattcctcaactcttgattttaactatcatatgatttaactttgtgattgaataat	524							

Db 2685 AAGATATCCATCTGTAAAAAAGCTTTTCAGTTCATCAATGCAACCATTAATATAC 2744
OY 2710 cctcaagacaagccgaatgagcgctaaagcgaacactcactcgcgcgaacctaat 2769
Db 2745 CCTCAGGATTAACCCGAGTAAGTGCATAATGATACATTCACATTTCAAAATTTGAAT 2804
OY 2770 aactcactgaaactgtctgagaaactcctgctcctcgcaagaataaagacatctacaa 2829
Db 2805 AATAGTTTAAATTTATTTGAAAAACATATGAGCTTTGGAAAAACAAATATATACCA 2864
OY 2830 gaactatttgacaagaatcgctcggaagactctcagagaagatctagaagacagcgac 2889
Db 2865 GAATTAATAGTCAAAAAGTAGTAAACCTTTTATGAAAAGATATTAATAATAGTGAT 2924
OY 2890 aactctataagagagcgctcactactcgtgaaactcgaagccgagatctactc 2949
Db 2925 ACATTTTAAATGAATCTTTTACAAATTTTGTAAATCTAAAGCTGATATATTAATTA 2984
OY 2950 cttaacgatgaaactcaacgttaagaagctggaagagacatcaataagctgaagaagaca 3009
Db 2985 TTGATGATGATCAAAAAGAGAAATAGAAAGAAATATTAATTAATAAAAAAAT 3044
OY 3010 ctgcaactgagctcgcaactcgtaacaagaagtaacaactgaaactggaagactctcgac 3069
Db 3045 TTACAGTTTATCATTTGATTTATATTAATTAATAATTAATAATTAATAATTAATTAAT 3104
OY 3070 aagaagaagacgctcgcaagatataagatcgacatcaagaatltgactctgctcaagag 3129
Db 3105 AAAAAAGAAACGTTGGTAAATTAATAATGAATAATTAATAAAACCTTATTAATAAGAA 3164
OY 3130 cagcttgaaagcaactcaactcactgaaacaatccgaacacgctcagagactctca 3189
Db 3165 CATTAATGATCAAAATTTGATTTCACTTAATTAACCCAAACATGATTAACAAAACCTTTCT 3224
OY 3190 gtgtcctcaacaagaagaagccgagatcgccgagacagaaacactctgagaac 3249
Db 3225 GTTTCTTTTAAACAAAAAAGAGCTGAAATAGCAAACTGAAACCATTTAGAAAC 3284
OY 3250 accaagatctctcctcaaacactcaagaagccctgcacgttatataatgagagctctc 3309
Db 3285 AAAAAAATTTTAAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAAT 3344
OY 3310 cctcgaagacactcctcggaagagacatccagacgagatcaactcgcgaaccccgag 3369
Db 3345 CCATTTAAAAAAGCTTTAGTGAAGATCAATTCAAACAGAAATATTAATCCAGTTTAGAA 3404
OY 3370 aactcgaagctcgtctcgaactcggaagcgaagctggaagacacactggaagag 3429
Db 3405 AACTTTAAAGTATTAAGTAAATTTAGAGGAAATTAAGATTAATTAATTAATTAAGAAAG 3464
OY 3430 aagaagctcagctactcctcctgagacgtcactcactcgtcgcgcgagctcaagaagc 3489
Db 3465 AAAAAATTTATCTACTTATCAAGTGAATTAATTAATTAATTAATTAATTAATTAATTA 3524
OY 3490 attaagaacaaagaactacacgcgcaatagcccaagcgagaatcaatacagaagctgaatac 3549
Db 3525 ATAAAAAATTAATAATTAACAGTAATTCCTCAAGTGAAGATTAATTAATTAATTAATTA 3584
OY 3550 gcaactggaactctcaagaagctcctcgctgaagaagacagatgctcgcaactgtgctc 3609
Db 3585 GCATTAATGATCTTACAAAAAATTTCTCCGAGAAAGACGATGTTCACACCTTTGAAT 3644
OY 3610 gaattgctcgcagacactggaagcgtctcaactaagaagccgcaactcactcactgctc 3669
Db 3645 GAAATGGATCCGACACATTAACAAAGTCAACCAAAAGAAACGACATCACTCATGTA 3704
OY 3670 ggaagcgagctcaatacataltaaccacatctcgaagcgtcgaagatgaggtctgactgctc 3729
Db 3705 GGAGCAGAGTCTAACACATATCAACATCAACAAATGTGATGATGAAGTAAATGACGTA 3764
OY 3730 atcatgtgctactactcgcgagagcggaagacactcagatgacccgcgcgaagtgtc 3789
Db 3765 ATCATAGTACCTATTTTGGAGAAATCCGAAGAAAGATTAATGATGATTTAGGACAGTACGTA 3824

OY 3790 accggtgagcgtcactcctcgcgtgattgaataactgtctccaaatcgagaagca 3849
Db 3825 ACAGGAGAGCGATACCTCTCCGTAATTTGATTAATCTTTCTTAATTAATTAATTAATTA 3884
OY 3850 tacgaagctcctatctgaaacctcctgcaagcgctcctataggtctctcgaagaagcgtg 3909
Db 3885 TATGAGCTTTTATTTAAACCTTTAGCAGGTGTTATAGAAGTTTAAAAAAACATTA 3944
OY 3910 gagaataacgtgactcctcaatgtcaagctgaaagacatctcgaagccgcttaac 3969
Db 3945 GAAATTAACGTTATGACATTTATGTTAATGTTAAGATTAATTTAAATTCACGATTTAAT 4004
OY 3970 aagaagaagaatcttcaagaagcgttggaagcgactgactcccatataagaagccgac 4029
Db 4005 AAGCGTAAATTTCAAAATGTTTGAATAGATTAATTAATTCATTAATAAGATTTAAACA 4064
OY 4030 tcccttaactacgtgtcgaagaagccatcaagtlccctcaataaagaagaagagataa 4089
Db 4065 TCAAGTAATTTATGTTGTCAAAGATCCATTAATTTCTTAATAAAGAAAGAGATTA 4124
OY 4090 tctcgtcgtatcaactatatacaagactccatcgacacagatataaattcgctaat 4149
Db 4125 TTCTTAACGAGTTATTAATTAATTAATTAAGATTCATTAAGATTAATAATTTTGCAAAT 4184
OY 4150 gatgtgctggttattcaagaagctccgagcgagaataaataaagctcgactcact 4209
Db 4185 GATGTTCTTGATTAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4244
OY 4210 aaaaagatalcaacgataaagaagcgagaatgaaatatactcctcctcgtataac 4269
Db 4245 AAAAAATATATACGACGACAAACAGTGAAATGAGAAATACCTTCCCTTTTAAACAT 4304
OY 4270 atcgaaacccgttacaagaagctgaaagcaaaactcgacactcgtgtaactcgaag 4329
Db 4305 ATTGAGACCTTTATTAATAACGATTAATGAATAATTAATTTTGTATTAATTAATTA 4364
OY 4330 gccaaagctcctcaactatacttaagagaagagcaatgtgaaagttaaaatcaagaagc 4389
Db 4365 GCAAAAGTTCTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4424
OY 4390 aactaccccaaaacatcccaagaagcgtgcagatttcaagaagaataaatacaactcgtc 4449
Db 4425 AATTAATTAATAACATTAATTAACATTAATTAATTAATTAATTAATTAATTAATTAAT 4484
OY 4450 ggaattgcagactcgtctcgaattataaacaacaactcctcgaagctcgtcgtc 4509
Db 4485 GGAATTTGCTGATTTATCAACAGATTAATTAATTAATTAATTAATTAATTAATTAAT 4544
OY 4510 acttgcatgtgtcgaagaagcgtccaaagacagtgctggaagacactcgtcgaagcaac 4569
Db 4545 ACAGGATGCTTTGAAATCTTGTAAACCGTTTATTAATTAATTAATTAATTAATTAATTA 4604
OY 4570 ctgcagggatgctgaacatcccaagcacaatcgttgaagaagaagtgccccaagat 4629
Db 4605 TTGCAAGGATGTTAAACATTAACATTAACATTAACATTAACATTAACATTAACATTA 4664
OY 4630 agcgcgtcttcaagcactcgaagcgtggaagcgtgcaagtgctcctcgaactaca 4689
Db 4665 TCTGATGTTTCAAGCATTTAGATGAAGAGAAATGTTAATGTTAATTAATTAACAA 4724
OY 4690 caagaagagataaagtgctggaagaacccaacccactcgtcgaagaagaagaagcg 4749
Db 4725 CAAGAGGATGATTAATGTTGAAATCTTGTAAACCGTTTATTAATTAATTAATTAATTA 4784
OY 4750 tgtgacgccaatgctlaaatgcaacgagaagcaagcgtcctcaagcaagaagaatacaca 4809
Db 4785 TGTGATGCGAGTCCAAATGTAACGAAGAGATTCAGTAGCAAGGAAAGAAATTCACA 4844
OY 4810 tgcgagtgtaacgaagccgaactcctacactcctcgaaggaatttctgctcagctc 4869
Db 4845 TGTGATGTAACCAACGATTTATCCACTTTCGATGAGGATTAATTTTCGACAGTCTCT 4904

[illegible]

RESULT	9			
LOCUS	AF062349	5271 bp	DNA	INV 09-FEB-2001
DEFINITION	Plasmodium falciparum strain HN2 merozoite surface protein 1 precursor (msp1) gene, complete cds.			
ACCESSION	AF062349			
VERSION	AF062349.1	GI:3859844		
KEYWORDS				
SOURCE	malaria parasite P. falciparum.			
ORGANISM	Plasmodium falciparum			
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
AUTHORS	1 (bases 1 to 5271)			
TITLE	Jiang, G., Liu, R., Daubenberger, C.A. and Pluschke, G.			
	Sequence analysis of the MSP 1 gene of Plasmodium falciparum from			
	Hainan, China			
JOURNAL	Chung, Kuo Chi Sheng Chung Hsueh Yu Chi Sheng Chung Ping Tsa Chih 17			
	(5), 294-297 (1999)			
REFERENCE	2 (bases 1 to 5271)			
AUTHORS	Jiang, G., Liu, R., Daubenberger, C.A. and Pluschke, G.			
TITLE	Direct Submission			
JOURNAL	Submitted (30-APR-1998) Molecular Immunology, Swiss Tropical			
	Institute, Socinstrasse 57, Basel CH-4002, Switzerland			
FEATURES	Location/Qualifiers			
source	1..5271			

```

"organism"="Plasmodium falciparum"
"strain"="HN2"
/db_xref="taxon:5833"
/chromosome="9"
/country="China"
66. .>5180
gene="mspi1"
product="merozoite surface protein 1 precursor"
66. .>5180
gene="mspi1"
gene="mspi1"
note="binds spectrin"
/codon_start=1
/product="merozoite surface protein 1 precursor"
/protein_id="AAC72885.1"
/db_xref="GI:3859845"
/translation="KKIIFPISFLPFIINTQGVTHESYQGLVAKTALDAVLVLTGSL
LFQKKRWLNBEETSGTAVTSTPGSGSGGSAVSGSGGSAVSGSAVASVAVSG
GGSGSRNRTPSPDSNDAKSVADLKHRYQNLFTIKELKYEELDLTHMMLTLCN
IHGEVYLLIDGIEYELIKLRFNYVYLLPAKLINDACANSQIIPFNKIANEALDVLK
KIVFERPDLNIDKDWGMEYIKKNKTTIAPINIEBESKTIIDONKADNEGKKK
KLYQVQNYFNKQLOEKNHLSYERKIDPLTKENKIKLKLIDINEIKNPPRANG
NTPNLDLNNKKNIIEBEKEIKELAKTKRNIDSFLPDLELEYLRKKNKVDVTPSS
ODPRTSVQIPVYPYPGVILYPLVDIHNLSLADNKNNSGDLMLNPRTKINIKLIT
NENKERKIFINNIKNIDLEEKNIHNRKEDNKLILEYEKSKYELIELEFEKPEFNKK
NDEKQVIEFARTYANDEKORYNKKFSSNSNVYQVKALSLYEIDYSLRGKIS
NEKFNHYTKIETDKLEIKLEEKSENNKILEKNKGLTHASLSEDIYKQVQY
OKVLLIKIEDLRKLEFLKNNQAKDSIHVPNITRYQNKREPIYLVLRKVDKAKEF
IPKYNDAKKEEAVLSSITQPLVASETTEGGSHHTLSQSETEVTEVTELEFEFV
GHATTYVITLPTQSPREPVEYVENSIEHKSNDNSQALTKRYVYKLKIDFLKYSJC
KHYLVNNSMPOKLEEVNLPPEEKNEKDCDPLDLNINNNIPKAYSLYQSMNND
LOHLEFELYQKMIYVYIKLKEENHKLLEBQKQITGTSSTSPGNTVYNTQASAHN
SNSQNOQSNASTNTQNGYAVSSGPAVVESEHDPVLVLVLSNLDKGLVSLNNGKTK
VNPNTLITTENEKEYEYENLKNNDYFENDIKQFVKSNSKYVIGLETKQNALDEIKR
KLKDTLQCSFDDLYNRYKLKLDRLFNNKKELEGDKMOIKLTLTKLEQLESKLSINP
HKLQNDVSEFNKKKEAEILEENTLENTLKHYGLVAYGSGSEPIKTSLESTQ
TEDNYANLEKPVLSIKQKLDNLDHLGKKKLSFSSGLHKLITTELKEYIKNNYCN
TSPSENNKYNELAKLEENFLPEKATVTVVPPQDVPSPLSVAVSGSGSETEETDI
ITSGSLVLELQGVQSNQDEDDSLVLPPIRESDBNDEYLDQVYVGAELVYMDNI

```

LSGEENYEIVYIAKPLKGLVGRSLKJKEJENIIFNNLNLINLSRKKRRKRYPLDIVLES
 DIMOFKJISSENYEYIEDSPKLNSEKQENKILSYKKYKESEVEDIPRTAIGISYEVY
 LAKKDLESIKKVIKKEKEEPSPPTPTSPKATDOKRESEFLPTATIGIELNLY
 YAKKDLESIKKVIKKEKEEPSPPTPTSPKATDOKRESEFLPTATIGIELNLY
 NDTKIDDKMLKGLSLSTGLVONPNPNTIISKLEIGKQDÖPMLN SÖHÖCKARQÖPEENSGKR
 HIDEHECEKCLCNTKÖEGDCKVCEPNPEPTCENNNQGGADADAKCTEEDSGSNKGRITCEC
 TKPSEYLFELGIRFSSSSNFGJISFLLLMLILLYLSFI

BASE COUNT 2313 a 687 c 690 g 1581 t
 ORIGIN

Query Match	28.4%	Score 1400.6	DB 5	length 5271
Best Local Similarity	57.3%	Pred. No. 2.9e-288		
Matches 2944	Conservative	0	Mismatches 1974	Indels 219
				Gaps 14
Qy	10	atgaatacatcttctccctcgtcatcttctgttttctatcatcaactcaactcagctg	69	
Db	66	ATGAAGATCATATTCCTTTTATGTTCATTTCTTTTATTTATTAATAACCAATGTGT	125	
Qy	70	accacgaatccctacagagcgtgttaagaactggaagcttggagaatgcccgtc	129	
Db	126	ACACATGAAAGCTTTATCAAGAAGACTGTCAAAAACATGAGACGTTTAAGAAGATGAGATATG	185	
Qy	130	accgaatcacgcctcttccagaaagaagaatgtgtgcgaatggaaggagagagccg	189	
Db	186	ACAGCTTATAGTTTATTTTCANAAGGAAAAATGGTATTAAATGANGAGACAGTGTGACACA	245	
Qy	190	gcgcgtacaacacagcacaccgcgtctctaaggctctgtgtcgaagcgtgtccggtgg	249	
Db	246	GCTGTTAACAATGATGACACCTGTGTTCAAAAGGTTTCAGTGTGCTCAGTTCTTCAGGTGT	305	
Qy	250	tctgtgacctcttgggggtctcgctgcgcccgggcagcgtgcatcaagtggtctcag	309	
Db	306	TCAGGTGGTTACAGGTGCTCAGTGTCTTCAGTGTCTTCAGTGTCTTCAGTGTCTTCAGTGT	365	
Qy	310	gcaagcgcgcgtctcccggaacagctcgaagaaccacatctatgcaactctagcagatcc	369	
Db	366	GCTTCAGGTGTGTTACAGTGTATTTCAAGACGTACAAATCCTTCAGATTAATTCAGATGTCA	425	
Qy	370	gaagcgaagctctacgcgcgaccccaagcagcagtgagaaactctctcactatacag	429	
Db	426	GATGCTAATATCTTACGCTGATTTTAAACATGTAGCTTCAAAATTTACTTCTTCATTTAA	485	
Qy	430	gagctgaagtcaccaagctgttgcagctcactatataatgctgcaacgttgtgataac	489	
Db	486	GAACTCAATATCCCGAAGCTCTGTGATTTTAAACAAATCATATGTTTAACTTTGTGATAT	545	
Qy	490	atcaatgcctcaaatatctatctatgacggtctcgaagaagatcaatgaactcctg	549	
Db	546	ATTCATGTTTCAAAATATTTAATGATGATGATGATGAGAAATTAATGATTTATTTATTA	605	
Qy	550	ttgaaattctactctgcagctctgaagggcacaactgaaatgacgttttcgcgaatgaact	609	
Db	606	TTAAATCTTTTATTTATGATTTATTTAAGACCCAAATTAATGATGATGATGATGATGAT	665	
Qy	610	tgctcaatctcatcaatcttgaagatcgaagcgaagcaggtctggaagcttgaagaagt	669	
Db	666	TGTCAAAATCCTTTCAATCTTAAATTTGTGTCAAAATGATTTAGACGTACTTAAAAAATTT	725	
Qy	670	gtctctggaatctcgaagcctctcgaacacatcaagaacaaatgttggagaagaatg	729	
Db	726	GTTGTTGCGATTTAAGAAAACCTATGACAAATTTTAAAGATATTTGAGGAAAAATGGAAGAT	785	
Qy	730	tatatataaagaataagaagacatctgagaacatctaagcgtctgcatctgagaatccaaa	789	
Db	786	TGATCTTAAAAAATAAATAAACACATACCAATTTAAATGATTAATTAATTAATTAATTAAT	845	
Qy	790	aagccatcagacaaataaagaatgcaacccaaggaaggaagaagaagaagtctgacag	849	
Db	846	AAAACATTTGATCAAAATTAAGATGCGATATATGAGACAGGAAAAAATAATTTTACCA	905	
Qy	850	gccagtaagcagctgtccatctatacaacaacagcttgaagaagccataaactcatcagc	909	

[illegible]

OY	2959	gaatctaaagctaaagacgtggaagaggaatcaataaagctgaagaagacactgcaactg	3018
Db	3120	GAACACCAAAAAATGCAATTAAAGATGAATTTAAAAATTTAAACATCTTTACAGTA	3179
OY	3019	agcttcgcgcctgtacaaacgaatcaaacctgaaactcgagagactcttcgcaagaagag	3078
Db	3180	TCATTGGATTTTATATTAATTAATTAATTTAAATTTGATNGATTTATTTAATNAGAAAAA	3239
OY	3079	acagtcgcgaagctaaagatcgatcaagatctgactctgctcaagagcagcttgaa	3138
Db	3240	GAACCTGGCCAAAGCAAAATTCGAATTTAAAAAACTTACTTTATTTAAAGAACAAATTAGA	3299
OY	3139	agcaaacctcaactcactggaacaatcgaaacacgctacgtgcagaactctcagtgcttc	3188
Db	3300	TCAAATTTGAATTTCACTTAATTAACCCACATTAATGTATTCCAAACCTTTCTGTTCTTT	3359
OY	3199	aacaagaagaaggaagcccgagatgcgcgagacgaagaacactcggaagaacccaagtt	3258
Db	3360	AAACAAAAAAGAGAGCTGAATTAAGCAAAACGGAACACATTTGAAAAACCAAAATA	3419
OY	3259	cttcctcaacactcaaaagccctgcgtaagctatcataatgagcgagctctcctctgag	3318
Db	3420	TTATTTGAACATTTTAAGAGCTTTGTAATATTTAATGATGGATCATCTTCATTAAA	3479
OY	3319	actctctcggagagagcatccagaccgagataactacgacgacgcctcgagaactcaag	3378
Db	3480	ACTTTAAGTGAAGTATCAATTTCAACAGAGATTAATTTGCCAATTTGAAAAATTTTGA	3539
OY	3379	gtccggtctaaagctcgaagcgaagctggaaggaacactgaaactcgagagagaagaagtc	3438
Db	3540	GTATTAAGTAAATAGATGGAAAACTCAATGATTAATTTCAATTTGAGAAAGAAAAATA	3599
OY	3439	agctaacctctcgaagcagctgcaactcagatgcgcgagctcaagaagctcaatgaac	3498
Db	3600	TCTTTCTTATCAAGTGGATTACATCAATTTAATTAAGATTTAAAGAGTATATAAAAT	3659
OY	3499	aagaactacacccggaatccccaagcgaagaataatacagacgtgataacgcactgaa	3558
Db	3660	AAAAATTAATACAGTAAATTTCTCCAGAGGAAAAATTAAGAAAGTTAACGAGCTTTAAA	3719
OY	3559	tcttaagaagcttcctcgctcgaaggaagaagatgctgcacatgctgctgctgc-----	3608
Db	3720	TCATTACGAAAAATTTTCTCCAGAAAGCAAAAGTTACAAAGTTGTAACTCCACCTCAACA	3779
OY	3609	-----tgaaatcgtcctcgacacacgcggagcgagcttcaaccctaa	3647
Db	3780	GATGAACCTCATCTCCATTAATCTGTAAGGGTAAGTGAAGTTACAGATCCCAAAAAGAA	3839
OY	3648	gaag-----cctgacatctactcatgctcgagccgagctcaataataaccaactc	3699
Db	3840	GAACACCAAAATACCAACCTTCAGCTCTTTAATTAACGAATTTCAACAAGTAGTACAAATTA	3899
OY	3700	cagaagctcgagatgagctgcgaatgcgcacatcatgtgcctactcttcggagagagag	3759
Db	3900	CAAAATTAATGACGAGAGAGTGAATTCCTTAGTTGTAATCCCATTTTGGAGAAATCCGAA	3959
OY	3760	gaggaactcagatgaaccccgacggcagctgctcacocggctgaaagctcactccctgatt	3819
Db	3960	GATTAATGCGAATTTTAATGATCAAGTAAGTAACGTGGAGAAAGC---AATATCTGTCAAATG	4016
OY	3820	gataaactctgtcccaaaatcgagaacgaatacgaagtgctctatctcgaacactctgca	3879
Db	4017	GATTAATTTCCCTTCAGATTTGAAATTAATTAATGATTTAATTTAATAAAGCTTTAGCT	4076
OY	3880	ggcgctaatgagctctcgaagaacgcctggaggaataagcgtgagactcttaagtgtaac	3939
Db	4077	GGAGTATATGAAAGCTTTAAAAAACAATTTGAAAAAACAATTAATTAATTTAAAT	4136
OY	3940	gtgaagacactctggaacgcgcctctaaatgaagaaatctcaagaagcgtcttgag	3999
Db	4137	TTGAACGATATCTTAATTAATTTACGCTCTTAAGAAACGAAAAATATTTCTTGATGATTTAGAA	4196
OY	4000	agcgacttgatccctataaagaactgcactcctctcaactaagtgctgcaagaagccaatc	4059

Db	4197	TCGTATTTAAAGCAATTTTAAACATATATTCCTCAATATGAATACATATTTATGAAATTCATT	4256
Oy	4060	aagttccccaataaagagaaagaggaataattctgtcagttgaacaactaacaagac	41199
Db	4257	AAATTATTGATTCAGAACAAAAAACACACTTTTAAAAAGTTACAAATATTTAAAGAA	43168
Oy	4120	tcacatgcacacgatatacaatttcgttaatgatgtgtcgtgggtatacaagaaccgtgac	41799
Db	4317	TCAGTAGAAAAATGATATTAATTTTGACACAGAAAGGTATAGTTATATACAAAAGTTTTTA	43766
Oy	4180	gaaataatacaagltctgacctgacctatataaagatatatacaagaataa-----	42300
Db	4377	GGCAATATATAAGATGATTTTGAATCATTTAAAAAGTTATCAAGAAAGAAAGAGCTTC	44366
Oy	4231	-----aaagcgagat	42424
Db	4437	CCATATATACACACACACACACTCGCTCACACCAAAAAACAGACGACAAAAAGAGAA	44966
Oy	4243	gaaaaatactcgtcccttcgttaatacatatgaaccctgtacaagaacagtgaacgacaa	43022
Db	4497	ACTAAGTTCCTTCCATTTTAAACAACATGTGACACCTTATACATTAACTTAATTAATAA	45566
Oy	4303	atcgacctctcgttaattcacctcgtgagcgcaagtcctcaactactaactgaagaagc	43622
Db	4557	ATTGACGATTACTTAATTAATTAATTAAGGCAAAATTAACGATTTGTAATTTGAAAAAGAT	46166
Oy	4363	aatltgagaaatlaaatacaagagcgtgaactactactaaacaatccaagaacagctgca	44222
Db	4617	GAGGACACTGTTTAAATTAATTAACCTTAGTGATTTTAAAGCAATTCATGATACCAAAATTAAT	46766
Oy	4423	gattccaagaanaaatacaaatlctgcgtgaattgcagacctgtcactgaatatatacac	44822
Db	4677	CTTTTAAAAACACTAAGCATTCCTGAGCAATTTAAAAAFTGATTAATGATATATCGAAA	47366
Oy	4483	aacaaatctctgacacagttctcgtccactgtgcatagtgtlctogaanaacctcgcacaa	45422
Db	4737	AAAGATATGCTTGGCAAAATTAAGTAGTACAGGATTTAG-----TTCCAAATTTTCTTAATACA	47933
Oy	4543	gtgcgtgagcaatctgcctgcagcggaacctgcgaagcatgtaaaatctcccaagacaa	46022
Db	4794	ATAATATCAAAATTAATTTGAGAGGAAATTTCCAAAGATATGTTAAACATTTTCACAAACACAA	48533
Oy	4603	tgctgtaagaacacagtgccccacgaatagcgcgtgttcaggacatctggaacgacgaa	46622
Db	4854	TCCGTAAAAAACAATGCCAGAAATTTTCGATGTTCCAGACATTTAGATGAAAGAAA	49133
Oy	4663	gagtgcaagtgtctcctgnaactacaacaagaagaataagtgctgtgagaacccaac	47222
Db	4914	GAAATTAATATTTTATTAATTTACAAACAAAGAGGTGATTAATGCTTGAAAAATCCAAT	49733
Oy	4723	ccatccctcaatgaanaacatgacggtgtggaacgcgatgtgcaaatgacgaagaagac	47822
Db	4974	CCTACTTTTAACGAAAAATTAATGCTGGATGTGATGCAAGATGCCAAATTTGACGAAAGAAAT	50333
Oy	4783	agcggtcttaacggaanaaatacaatcgaatggaatgtgactaaagcccgacatccatcacct	48422
Db	5034	TCAGGTACACGGAAGAAATACATGTGATGATTAACACGATTTCTTATCCACTT	50933
Oy	4843	tttgaacggatlltttgctccacgcctcaatttcctgtgggcattcctcttcgcgtacatc	49022
Db	5094	TTTCAGTGGTATTTTTCACAGTTCTCTTAATCTTAAGTAATTCATCTTATTAATATCTC	51533
Oy	4903	atgcgtacatctgtaacgactcatcatatgatcatg	4939
Db	5154	ATGTTAATATTAATTAACAGTTTCAATTTAAAAAATGTAGG	5190

RESULT_10

AF062348

LOCUS

DEFINITION

AF062348

5243 bp

DNA

INV

09-FEB-2001

Plasmidium

falciiparum

strain HNI mezoizolite

surface protein 1

precursor (mspi) gene, complete cds.

ACCESSION AF062348.1 GI:3859842
KEYWORD malaria parasite P. falciparum.
SOURCE Plasmodium falciparum
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium
REFERENCE 1 (bases 1 to 5243)
AUTHORS Jiang, G., Liu, R., Daubenberger, C.A. and Pluschke, G.
TITLE Sequence analysis of the MSP 1 gene of Plasmodium falciparum from Hainan, China
JOURNAL Chung Kuo Chi Sheng Chung Hsueh Yu Chi Sheng Chung Ping Tsa Chih 17
REFERENCE 2 (bases 1 to 5243)
AUTHORS Jiang, G., Liu, R. Z., Daubenberger, C.A. and Pluschke, G.
TITLE Direct Submission
JOURNAL Submitted (30-APR-1998) Molecular Immunology, Swiss Tropical Institute, Sociustrasse 57, Basel CH-4002, Switzerland
FEATURES
source
1..5243
/organism="Plasmodium falciparum"
/strain="HNI"
/db_xref="taxon:5833"
/chromosome="9"
/country="China"
/gene="msp1"
/product="merozoite surface protein 1 precursor"
/note="binds spectrin"
/codon_start=1
/product="merozoite surface protein 1 precursor"
/protein_id="AAC72884.1"
/db_xref="GI:3859843"
translation "MKIFFLCSFLPIINTQCVTHESYQELVRLKLEADAVLVGYSL
LFQEKWVNLNGSTGTAFTTSTPGSGSVTSGSGSVASVAGSGSGSVAGSGNS
RRTSPNSSDSAKSYADTLKRVONTLFTTKEIKYPELPFLTHMLTCTHNGPKY
LIDGFEINELTYLNFYDILRAKINDACNSYCOIPENKIRANEDLVKTIYEGY
RKPLDINKDVGKMEYIKKRTTANINELBESKTIIDDKNADNEEGKKIKLYQV
NLFYLNQLOEHLNLSVLEKRIDTILKNNIKKLEDDIKITDAKELTGTGSPNP
LPEKKKEVEGHEKIKELAKTIFENIDSLTLELEYLRKKKQKVDVPKSDPT
KSVQIPKPYNGIYVPLPTDINHSLAADNDKSYGLAMPDREKNEKITDNKE
RKIFINNIKQIDLEKNNHTEKONKLLDYKSKDYELLEKEFEMENNNEDK
DYNDKISARTYVNEKQRYNNKSSSNSTYNNQKLKASLYEDLSLRGISGSL
NHYTLTGLEADIKLLEETIKSEENKILEKNEFGLHSANALEVSVLQVQKVL
LKKIEDLRKLEIFLKNQKDSLHPNITVPKPKPEYVILVLEKEVDKLEKEIPVY
KMLKEDQAVLSITQPLVASSETTEGSHSTLLOSGETEETEEYGHATTYITL
PTQSPPEKVVYENSIEHKSNDNSOALITVYLKDKDEFITSYICHKIIIVNS
MDOKLELVNLTPEKNEIKSDPLDILFNONTNPAYSLYDSMNDLQHLFELYO
KEMITYLHKLEENHIKKLEEQKITGTSSTSPGNTTVAOSATHSNONOONA
SSTWQNGVAVSSGPAAVEESHDPVLVLSINDKGLVSLINDKTKVNPPLTIST
EMEFYENILKNNDTYFNDDIKQFVNSKVTGLTJEOKNALDELKIKDLTQLSP
DLNKKYKLDRLFNKKKELGDKMOIKLTLKEOLESKLNNPNHNVONFSVF
NKKEAEIAETENTLENTKILKHYKGLVKNSSPLKTLSEVSIOTEDNVANLEK
FRVSKIDGRKNDMLHKKLSPLSKLHLITELKIKYTKNNKVTGSPENKKVY
EALKSYENLEPEAKVTVYVTPPODPVTPSPISVSYSSSGSTKETQIPTSGSLTEL
QOVSQNYDEDEDLVVLPFGESEDEDLVDVYVGEALVTMDNLTSFGEYEDY
IYLPAGVYASLKQKLEKNTITNLNINDILNLRKFKKQFVLEEDLQKFHIS
NEYIIEFSFKLNSQKNTLKSYYKIKSEYENDIKFQKESISYEYKVLAKYDLES
IKKVIKEKEFPSSPTTPSPAKTDEKRSKFLPFLNTITETLYNNLVNKTIDYLLN
LKAKINDCNKEDAHVKTIKLSDKLADKIDLFPKNDPEALIKLINDTKDMIG
KLSTGLYONFPNTISKILGKQFODMINISOHCKVKQCEPNSGCRHLDREBCKC
GLNKQEGDKVENPNPTCENNGSCDADACTEDSDSGNKKTKTCECTKPSYPLFD
GIFCSSNSFLQISPLILMLILVGSFI"

BASE COUNT 2303 a 677 c 686 g 1577 t
ORIGIN

Query Match 26.9%; Score 1331.2; DB 5; Length 5243;
Best Local Similarity 56.9%; Pred. No. 1.9e-273;
Matches 2919; Conservative 0; Mismatches 1978; Indels 231; Gaps 17;

QY 10 atgaataatcttctctctctctcattctcgtttttatcatcaatactcagtcg 69
DB 66 ATGAAGATCATATTTCTTTTATAGTTCATTCTTTTATATATAACCAATGTGA 125
QY 70 acccaagaatccatcagaagctggttaagaactggaagcttggaaatgcgcttc 129
DB 126 ACACATGAAGATTATCAAGAACTTGTCAAAAACATGAGAGCTTTTGAAGATGCACGATTTG 185
QY 130 accgatatacagctgttccagaagaagaatggtgcgaatgaaggagcagtcgacg 189
DB 186 ACAGGTATATGTTTATTTCAAAAAGAAAAAGATTAATAAGAGAACCACTGGAACA 245
QY 190 gccgttaacaccagcaccccggttctaagggtctgtcgtgtaggagtcgacgctg 249
DB 246 GCTGTACACAGTACAGTACACCTGTTGAGTGGTTAGTACTTCAAGTGTGTTCAAGTGTG 305
QY 250 tctgtgacctctgggggttcgctgcctccgcgcaagcgttgcatcagtgctcagtg 309
DB 306 TCAGTTGCTTCAGTTGCTTC-----AGTGTGTTCAAGTGTGCTCAGTT 347
QY 310 gcaagcgcggttcgcggaacagtcgaagaacacatccatcgaactctagcgatcc 369
DB 348 GCTTCAGGTGTTCAAGTATTCAGACGTACAAATCTTCAGATTAATTCAGTATTC 407
QY 370 gacgcaagtcctcagccagaccctcaagcacccgagtcgagaacatcctcactcaag 429
DB 408 GATGCTAATCTTACGCTGATTTAAACATAGAGTTCAAAATTTACTTGTCTCATTTAA 467
QY 430 gagctgaagtaaccacagctgttcgacctcaactaatcatatgacacgtgtgataac 489
DB 468 GAATCAAAATATCCGAACCTTTGATTAAACCATCATATGTATTAACCTTGTGATTAAT 527
QY 490 atctgcttcaaatatcgtatgacggttgcgaagatatagaactcctgtaacag 549
DB 528 ATTCATGTTTCAAAATTTAATGATGATGATGAGAAATTAAGAAATTAATTAATAA 587
QY 550 ttgaattctactcagctgtgaagggcaaacgaaatgacgttgcgcaactactac 609
DB 588 TTAACCTTTATATGATTATTAAGAGCAATTAATGATGATGATGATGATGATGAT 647
QY 610 tgcataatccatcaatttgaagatcagagcaagcagctgtgacgtatgaaagtg 669
DB 648 TGTCAAAATCTTCAAACTTAAATTCGTGCAAAATGAAATGAGCTTACTTAAAAAAT 707
QY 670 gttctcgatatacgaagcctctgacacatcaaggaacatgtyggaaagtgtgaag 729
DB 708 GTTTCGATATGAGAAACCATTAAGACATATTAAGATTAATGTAAGAAATGGAAGAT 767
QY 730 tatattaaagaataagaagacatcgagaaacattgaagcgtgatcgaagaatccaaa 789
DB 768 TACATTAATAAATAAATAAACAACCATAGCAATTAATAATGAATTAATGAAGAAAGTAA 827
QY 790 aagacacatagacaataataagaatgcaacccaagggagaaagaagaagtgtaccag 849
DB 828 AAAACAATGATCAAAATATGAAGATGAGATATGAAGAAAGAAAAAATAATATACAA 887
QY 850 gcccaagtaagcctgtccatctatacaaaacagcttgaagaagcccaataacctcagc 909
DB 888 GCTCATATAAATCTTTTATTTACATAATAACAAATTAACAAGAACCATTAATTAATGAC 947
QY 910 gtaacggagaagcgaatagacacccctcaagaagaatgaaatatacaagaactgtcgac 969
DB 948 GTTTTAGAATAAGATATGACACTTTAAAAAATAATGAATAATTAAGAAATTAATCTTGA 1007
QY 970 aagatataatgaatataaga---atcctccgcaagcaactctcggaacacccctcaacg 1026
DB 1008 GATATAGATATAAATTAATAACAGATGCCGAATAATCTACACTGGAAGTAAACCAATCT 1067
QY 1027 ctgctggaagaagaagaagata---gaggaagcaggaagaagatcaagaagatcgcc 1083
DB 1068 CTCCTGAGAAATTAAGAAATAAGAACTGAGGAGCACAGAAAGAAATAATGAAGAAATTTGCC 1127

1084 aaacattaaagtcacacatagattctcttactgactcccttgagctggagctaac 1143
1128 AAACTATTAAATTTAACTTTGATGTTTATTACTGATCCACTTCATTAATTAATTTT 1187
1144 ttgagagaagaataaagataatagacatccgcgaagtcga-----gacaaag 1194
1188 TTAAGAGAAAAAATAAATAAGTTGATGATACCTTAATACAAAGATCTACGAATCT 1247
1195 gaataacccgaaccgaataatcccaatgggtgagcttaccctctgctctataagct 1254
1248 GTTCAAAATACCAAAAGTCTCTTATCCAAATGGTATTGTATATCTTTACCTCTCATCTG 1307
1255 atcaacaa-----cgcttcaacgagcccaatagcttcggtgacttgatacccc 1305
1308 ATTCAATATTCTATTGCTGCATATATGATTAATAATTCATATGCTGATTTAATGAATCT 1367
1306 ttgcattacgaagaaccccttaagaatctacacagaacatgagagaagaagttt 1365
1368 GATACTAAAGAAAAAATTATGAAAAAATTATTACAGATTAATAGGAAGAAAAATATTTC 1427
1366 atcaacgaatcaagaagagatcaaaatltgagaagaagaatltgagatgacaaga 1425
1428 ATTAAATACATTAAATAACAAATGATTTAGAGAAAAAATCAATTAATCACACAAAAGTA 1487
1426 agtaccgaagaccgagcaaaagcttaaacgatatcaactaaagatgtgaaaaagctgtg 1485
1488 CAAAATTAATAATTTCTGATGATTAATGAAGTCAAAAGATTTATGAAATTAATTTACTT 1547
1486 aagagatcatgattccaaatcaacatcaacatcgaccctgacaacttcgagaagatg 1545
1548 GAAAAATTTTATGAATATGAATTTATATATTTTATGACAAAGATGCTGATATTAATA 1607
1546 atgggaaacggtactcttacaagaatgtgagaagaacgacacacataatctcactc 1605
1608 TTCACTGCAAGATATACATATATGTTGAAAAACAAGATTAATTAATTAATTTTCATCTC 1667
1606 tatgagaattcttaacataatcttgaagaagctcaccagaagctcttaagataatgagag 1665
1668 TCTAATTAATTTCTGTATTAATGTTCAAAAGATTTAAAAAGGCTTTTCAATCTTGAGAT 1727
1666 tattcttcgcggaacatcgtgtgtgagaagaacataagatltacaagaatcctcataag 1725
1728 TATTTTAAAGAAAAAGAAATTTCTGAAAAAGATTTTAAATCATTAATTAATCTTGAAAACT 1787
1726 aagatcgaagaagatcgagacgctgtgtgagaacatcaagaagaagaagaagaagctgtg 1785
1788 GGCCTCGAAGCTGATATTAATAAATTTAACAGAAATTAAGAAAGTAGTGAACAAAATTT 1847
1786 ttgagaagaagatcacaaaagaacgaataaaccagatgagaagaatccttgaggtctcc 1845
1848 CTAGAAAAAATTTTAAAGACTAATCAGCAATG---CTTCTTACAGATATCT 1904
1846 gatattgttaagaatcccaagtgagaggtgtcctcctcctcctcctcctcctcctcctc 1905
1905 GATATTGTAATTAATTAACAGTACAAAAGTTTATTATTTAAAAAATATGAAACCTTAATA 1964
1906 aagatcgaacatctcgaagaagctgtgagttaaacataatataatctgtccgaatagt 1965
1965 AAGATAGAAATTTTAAAAAATGCACACATTAATAAAGATAGTATTCATGTACCAATATAT 2024
1966 tataagcagaagaataagcagaacatactactcctcctcctcctcctcctcctcctcctc 2025
2025 TATAAACCAAAATTAATAACACACATATATTATTAATGTATTAATAAAGACAGATGAT 2084
2026 aaactcgaagaagttcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 2085
2085 AAATTAAGAAATTTTAAACCAAAAGTAAAGACATGTTAAA---GAAAGAACAAAGCTGTCT 2142
2086 aaactcgaagaagttcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 2145
2143 TATCAAGATTAATACACACCTTTAGTTGACGACGAAGGAAACAACTGAAGAT-GGGGCTCAC 2201
2146 gctacacacaaagcccgagacaaagcggltcagcttcgaagcgatagcgtgcaagct 2205

2202 TCCACACACACATTAATCCCAATCAGAGAAACAGAAAGAAACAGAAACATAGACAC 2261
2206 caagcagaagaagcagaagcagaagcctcagctcagctcagctcagctcagctcagctcag 2265
2262 ACACACACGCTTAACATTAATCATTACCAACACACACACACACACACACACACACACAC 2321
2266 gctcagaagtcctcaccacacagcctcctcctcctcctcctcctcctcctcctcctcctc 2325
2322 GTTGTGAAAAATTTCAATTAACATTAAGACTTAATGACAAATTCACAAAGCTTGACAAACA 2381
2326 gaactccttgagaagcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 2385
2382 GTTATCTTAAGAAATTTAGATGAATTTTAACTAAATCATATATATGCTATTAATATAT 2441
2386 ctcgctctcaccagaactatgaaagagaagatltcctaaagttacaagaataacaaagaa 2445
2442 TTAGATCAAACTCTAGATGACCAAAATTTATAGAGATATATATCTTCTCCAGAA 2501
2446 gagagagaataacgtccctcctcctcctcctcctcctcctcctcctcctcctcctcctc 2505
2502 GAAAAAATGAATTAATTAATCATATGATGATCCATTGATTTATTTAATTAATTAATAATAC 2561
2506 attccggtatgtatctatgtctcctcctcctcctcctcctcctcctcctcctcctcctc 2565
2562 ATACCTGCTATGATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 2621
2566 gagatataatgagaagaagatgtctcgaacgtctataaactcaagaacagaagaat 2625
2622 GATTAATATCAAAAGAAATGATTTATTTATTAATTAATTAATTAATTAATTAATTAATTA 2681
2626 aagaacctctcgaag----- 2641
2682 AAAAAATTTTATGAGAGAGCAAAACAAATTAATGCAATCTATGATGATGATGATGATGAT 2741
2642 -----aagcgaagaagctccacactcgtttaaactcctcctcctcctcctcctcctc 2696
2742 AATACACCGTAAATTAATCTGCTCAATCCGCACTCAAGATTAATCCAAACCAACAAATCA 2801
2697 actgctctcacaactcaagaac-----gcccggaagt 2730
2802 AATGCACTCTTACCAATTAACCAAAATGCTGATGATGATGATGATGATGATGATGATGATG 2861
2731 agcgtcgaagaagcagaacacactcctcctcctcctcctcctcctcctcctcctcctcctc 2790
2862 GAAAGAAAGTCATGATCCCTTAACAGATTTGCTATTAAGAACATTTGAAAGCTATGTT 2921
2791 aacatcctgtctcctcgaagaataagaacat---ctaccaaagaactltgacaga 2847
2922 ACTCTCTTAATCTTGAAATTAACCTAAAGTACCTTAATCCATTTAACTTTCTACACAA 2981
2848 tctgcgaagaacttctcgaagaagatcctcgaagaagcagaacatcctcctcctcctcctc 2907
2982 GAGATGAAAAAATTTTATGAGATATTTTAAAAAATTAATGATATCTTATTAATATGAT 3041
2908 ttcactaactcgttgaaatcctaaagccggtgatatcaactcctcctcctcctcctcctcctc 2967
3042 ATCAAAACATTTGTAATTAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3098
2968 cgtagaagcttgaagaagacataaagctgaagaagacatcgaactgagctcgaac 3027
3099 AAAAAATGCAATTAATGATGAATTAATAAATTAATAAAGATTTTACAGTTATCATTTGAT 3158
3028 cgtcacaagaatcacaactcgaagcctcctcgaagaagcagaagaagcagctcgc 3087
3159 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3218
3088 aagtaagaatcagaatcaagaagttgagctcctcgaagaagcagaagcagaagcagaac 3147
3219 CAAGCAAAATGCAATTAATTAATAAATTAATTAATTAATTAATTAATTAATTAATTAAT 3278
3148 aactcctgaaacatcgaagaacagctcctcgaagaactcctcgtgtcctcacaagaag 3207

Db 3279 AATTCATTAATACCCACATTAATGATTACAAAACCTTTCTGTTCTTTTAACAAA 3338
 Qy 3208 aagaagccgagatcgccgagacagagacactctggaacacacagatctctcaaa 3267
 Db 3339 AAGAAGCTGAAATGCGAACTGAAACACATTAGAAACACAAAATATTATTGAAA 3398
 Qy 3268 caclcaaaagcctcgtcaaglatataatggcgagctctcctcctgaagcctctcc 3327
 Db 3399 CATTTAAAGGACTGTTAAATATTATATGATGTAATCATCTCCATTAAACCTTTAAGT 3458
 Qy 3328 gaggagacatccagacagagataactacgcagcctccgaaactcaagtcctgctc 3387
 Db 3459 GAAGTATCAATTCACAAAGAAATATATGCAATTTAGAAAATTTAGACTATTAACT 3518
 Qy 3388 aagctcgaagcgaactgaagagacactgaacctgagagaaagaagctgcactcc 3447
 Db 3519 AATAATGATGGAACCTCAATGATTAATTTACATTTAGAAAACAAAATTAATCTTTA 3578
 Qy 3448 tctagcgactgcatcaactgcatcgccgagctcaaggaagtcatttaagacaagactac 3507
 Db 3579 TCAAGTGGATTACATCATTTATTAATGATTAATAAGAAAGTAAATAATTAATTAAT 3638
 Qy 3508 accggcaatagcccaagcagagataatcagacgctgaatacgcacgtgaactctacaag 3567
 Db 3639 ACAGGTATTTCTCCAAAGTAAATATTAAGAAAGTTAACGAAGCTTTAAATCTTACGAA 3698
 Qy 3568 aagtcctcgctgaagagacagatgctgcactgctgctgctc----- 3608
 Db 3699 AATTTTCTCCCAAGAAAGTAAACAGTTGTAATCTCCACCTCAACAGATGTAAT 3758
 Qy 3609 -----tgatctgctccgacacactgagcagctcgaactcaaga----- 3650
 Db 3759 CCATTCATTAATCTGTAAGGTAGTGTAGTTCAGATGTCACAAAAGAAACACAA 3818
 Qy 3651 --gcctgcatctactcatgctgagcgcagctcaataactcaacactctcaagaagctc 3708
 Db 3819 ATACCAATTCAGCTCTTTATTAACAAATTTACAAAGTGTGTAATCAACAAATTAAT 3878
 Qy 3709 gacgttgagctgcatgaagctcatctgctcctatcttcgagagagcagagagactac 3768
 Db 3879 GACGAGAGATGATTCCTTGAAGTATTAATCCATTTTGGAGAAATCCCAAGATTAATGC 3938
 Qy 3769 gatgaccccgccgagctgcatcagcgcgctgagctcctccctgctgaactcaact 3828
 Db 3939 GAATTTTATGATCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3995
 Qy 3829 ctgctcaaatcagagacagatcagagctgctcctgaactcctcctgagcagctcctat 3888
 Db 3996 CTCTCAGATTTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4055
 Qy 3889 aggtctcctcaagaaacagctgagagaaatcagctgagcctcctcaactgctcaagcag 3948
 Db 4056 AGAAGCTTAAAAAAACAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAA 4115
 Qy 3949 attcgaacagccgcttcaataagagaaattcgaagacgctctgagagcagctg 4008
 Db 4116 ATCTTAATTAATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4175
 Qy 4009 attcctcaataagacgctcctcctcaactcgtctgctcaagagcccatcactcctc 4068
 Db 4176 ATGCAATTTAAACATATATTCCTCAATGATGATGATGATGATGATGATGATGATGAT 4235
 Qy 4069 aataaag 4128
 Db 4236 AATTCAGAACAAAAACACATTTTAAAAAGTTTCAATATATTAATAAATCACTAGAA 4295
 Qy 4129 accgaatcaatctcgctcaatgctgctgagctctcaagaactcctgagcgaagatc 4188
 Db 4296 AATGATATTAATTTGCAACGAGAAAGTATGATGATGATGATGATGATGATGATGATGAT 4355
 Qy 4189 aagctgacactgactcattataaaglatatacaagataag----- 4230
 Db 4356 AAGGATGATTTAGATCAATTAATAAAGTTATCAAAAGAAAGAAAGATTCATCATCA 4415

Qy 4231 -----caagcgagagatgaataat 4251
 Db 4416 CCACCACACACACCTCCGACACGCAAAAACAGCAACAAAGAGAAAGTAAGTTC 4475
 Qy 4252 ctgcctcctcgtataaacaacagaaacctgtacaagacgtgaacgaacaaatcgacctc 4311
 Db 4476 CTTCATTTTAAACAAACATGAGACCTTATCAATTAATTAATTAATTAATTAATTAAT 4535
 Qy 4312 ttcgttaattcaccctgagagcgaagtlcccaactactactaagagagagagagagagag 4371
 Db 4536 TACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4595
 Qy 4372 gttaaatcaagagcgtgac 4431
 Db 4596 GTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4655
 Qy 4432 aaaaatacaattctcgtcgaatcgcagacacacacacacacacacacacacacacac 4491
 Db 4656 AACACTAACGACTTCGAACCAATTAATAAATTAATTAATTAATTAATTAATTAATTAAT 4715
 Qy 4492 ctgaccaaagttctgtcgcagctgcatgctgctgaacacacacacacacacacacac 4551
 Db 4716 CTTCGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4772
 Qy 4552 aatcgtcgcagcgaac 4611
 Db 4773 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4832
 Qy 4612 aacagctgccccagagataagcgcctgttcagacacacacacacacacacacacacac 4671
 Db 4833 AACAATGTCGAGAAATTTCTGATGTTTCAACATTTAGATGAAGAAACAAAGTAA 4892
 Qy 4672 tgtctcgtgaactcaacaagaagagagataagtgctgagagagagagagagagagag 4731
 Db 4893 TCTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4952
 Qy 4732 aatgaataacatgagcgcgtgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4791
 Db 4953 AACGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5012
 Qy 4792 aacggaagaataacacatgagctgagctgagcgcgcgcgcgcgcgcgcgcgcgcgcgc 4851
 Db 5013 AACGGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5072
 Qy 4852 atttctgctcagctcaattcctcgtgacatcctcctcctcctcctcctcctcctcctc 4911
 Db 5073 ATTTCTGCGAGTTCCTCAATCTTGAAGATATCATTTCTTAATTAATCAATGTTAATA 5132
 Qy 4912 ctgtacagcttcatcctaagatgagtg 4939
 Db 5133 TTATACAGTTTCAATTAATAAATGTTAGG 5160

RESULT 11
 AF218248
 LOCUS AF218248 5085 bp DNA INV 24-JAN-2000
 DEFINITION Plasmodium falciparum major merozoite surface antigen (gp195) gene,
 complete cds.
 ACCESSION AF218248
 VERSION AF218248.1 GI:6740099
 KEYWORDS
 SOURCE
 ORGANISM
 Plasmodium falciparum
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE
 1 (bases 1 to 5085)
 AUTHORS Shan,Z.X., Yu,X.B., Li,X.R., Ma,C.L. and Fang,J.M.
 TITLE Molecular cloning and sequence analysis of major merozoite surface
 antigen(gp195)gene of Plasmodium falciparum isolate FCCL/HN
 JOURNAL
 REFERENCE
 2 (bases 1 to 5085)
 AUTHORS Shan,Z.X.
 TITLE Direct Submission

OY	1366	atcaacgaatcaagaggaagatcaaatltggaaggaagaattcgagatgcagaaga	1425
Db	1363	ATTAAATTAACATTTAAAAACAATTTGATTTGAAGAAAAAATCATTAATCACACAAAGAA	14222
OY	1426	agttacgaagaccgcgcgcaaaagtlccaaagatatactcaagaglatagaagactgctg	1485
Db	1423	CAAAATTAATAAAATTTACTTGAAGATTTATGAAGAGTCAAAAAAGGATTTATGAAGAAATTTACTT	14822
OY	1486	aacgagatctatgatltccaatccaatcaaatatacatcgactgcaccacttcgagaagt	1545
Db	1483	GAATAATTTTATGAATAAGAAATTTTATTAATTAATTTTGACAAAAGATGTCGTATTAATAA	1542
OY	1546	atgggaacacggtactcttcaagaatgaggaactgcacacacataacttcctgcacc	1605
Db	1543	TTTCAGTCCAAAGATATACATTTAATGTTGAAAAACAAAAATATAATATAAATTTTTCATCC	16022
OY	1606	tatgaagattctgaagcctaactcttggaagaccaccaagactcttaagtataitgagac	1665
Db	1603	TCTAATTAATTTCTGATATPATTTGTTCAAAAATTTAAAAAGCTCTTTCATATCTTGAAAT	1662
OY	1666	tattctctcggaacatltgtctggaagaagaactaaglatlataagaatctcataag	1725
Db	1663	TATTCCTTTAAGAAAAAGCAATTTCTGAAAAAGATTTTAATCATTTATTAATCTTGAAACT	17222
OY	1726	aagtcgaaaaagagatcgagagccttgttggagaactataagaagatggaagacagt	1785
Db	1723	GGCTCCAAACCTGTATATAAATAAATTTACAGAAATATAAGATGAGAAAAACAAATTT	17822
OY	1786	tttgaagaagaatctacaagaagcgaanaataaacagatggaagaatcctgagctcc	1845
Db	1783	TTGAAAAAATTTTAAAGCATACACATTCACGAATG---CTTCCTTAGAAGTATCT	1839
OY	1846	gatatgttcaagtlccaagtgcagaagtgtcctccatgagacaagattgtgaactcaag	1905
Db	1840	GATATTTGTAAATTTACAAAGTCAAAAAGTTTATTAATTTAAAAAATGGAAGACTTAAGA	1899
OY	1906	aagactcaactcatctcgaagaacgtyggagttaaacataatacatgctgcgaatgt	1965
Db	1900	AAGATACAAATTTATTTTAAAAAATGCACAACTTAAGATGATTCATGTCACAAATTT	1959
OY	1966	tataagcagagataaagcaggaacacatactactcctatcgtatactcaagaagaatagac	2025
Db	1960	TATTAACACCAAAATTAACACGAGACCATTTATTTAATGTATTTAAAAAAGAGTGAT	2019
OY	2026	aaactgaagtgcttcgtgcccaagtcgagagcgtgatataagaagaagaagaacatt	2085
Db	2020	AAATTTAAAAAATTTTATACCAAAAAGTAAAAAGCATGTTAAA--GAAACACACAGCTGCT	2077
OY	2086	aaaaactgaagcagctcagataaactcgcgaccttcacagaagaagagataacgcgacag	2145
Db	2078	TATCAATGATTTACACAACCTTTAGTTGACGAAGCAAGCAACATGAAT--GGGGGCTAC	2136
OY	2146	gctaccccccgaagccgcgacaacagccggtttagctcttcgaagcgatagcgtgcaagct	2205
Db	2137	TCCACACACACATTTATCCCAATCGAGAGAAACAGAAACGAAGAAACAGTGGACAC	2196
OY	2206	caagcacaagaagacagaagcagagcacagccttcagctgcagctgccccgttccagagctaa	2265
Db	2197	ACAACAACGGTAAACATTAACATTAACCAACCAACCAACATCAACCAACCAAAAGAGTAAA	2265
OY	2266	gctcaagtgcctacaccaaccagctcctgtgtaataacaagaaccggaatgltcagaacactg	2325
Db	2257	GTTGTTTGAATAATTCATAGAACCTAAGAGTAAATGACAAATTCACAAGCTTGACAAAAACA	2316
OY	2326	gactacccttgaagactctataggttctcgyaatcaatccttaactctgcacaaatatatc	2385
Db	2317	GTTTATCTTAAGAAATTAAGATGAATTTTAACTTAATCTATATATGTGCATAAATATATTT	2376
OY	2386	ctcgctctccacagcactatgaaaggaagattcttaaacaglatagaataaccaagaa	2445
Db	2377	TTAATATCAACACTCTGTATTTGAGCCAAAAATTTATGAGAGTTATATAATCTTACCTCAGAA	2436

[illegible]

Db	3514	TCAGTGGATTTACATTCATTTTAATTACGATTTAAAGAGTATATAAAATATAAAATTTAT	3573
Oy	3508	accggcaatagcccaagcggagaataatatacagcgtgataacgcacgtgnaclttacaag	3567
Db	3574	ACAGGTAACTTCCACAGTGAATAATATAAAGAAAGTTAAGAGAGCTTTAAATCTTACGA	3633
Oy	3568	aagttccgccttggaaggaagatgtgcgaacgtgtgtc-----	3608
Db	3634	AATTTTCTCCAGAGCAAAAAGTTTACACAGTTGTAACTTCCACCTTACACCGATTTAACT	3693
Oy	3609	-----tgaaltcgtctcgacacacactcggagcagtlctcaactaaga-----	3650
Db	3694	CCATCTCCATTAATCTGTAAGGGTAAGGTGAAGTTCAGAGATCCACAAAAGAAACACAA	3753
Oy	3651	--gcttgatctactatgtcgcgagccggatccaatacaataccaacatctcaagcgtc	3708
Db	3754	ATACCAACTTCAGGCTCTTTATTTAAACGAATTTCAACAAGTGTACATCAACAAAAATTTAT	3813
Oy	3709	gaagatagtgctgaatgacgcgtacatctgtgcctatctcggcgagagcggaaggaactac	3766
Db	3814	GACGAGAAGAGTATNCTCTTAGTTGTATTCACCATTTTGGAGAAATCCGACATATATGAC	3873
Oy	3769	gaagactcggcgaagtgctacacggtgaggtcgtcaactccttcgltgaltgataacat	3828
Db	3874	GAATATTATAGTTCAGTAGTAGTACTGGAGGAGC---AATATCTGTACATGATGATATATC	3930
Oy	3829	ctgtccaaaatcggagacgaatacgaagtgctctatctbaaactctggcaggcgctcat	3886
Db	3931	CTTCTCAGGATTTGAAATGATATGATGTTATATATTTAAACCTTTTACTGGAAGTATAT	3990
Oy	3889	agttctcctcaagaaacagctcggagaatacgtgaltgaccltcaatgtgcaagtgaaagac	3948
Db	3991	AGAGCTTAAAAAACAATTTGAAAAAACAATTTACATTTAATTTAAATTTGAACGAT	4050
Oy	3949	attctgaaacgcgctttaaagaagaataatccaagaactcgtctggaagcgactgt	4008
Db	4051	ATCTTAATTTCCGCTTTTAAAGAAACGAAATATTTTGTAGATGATTTAAATCTGATTTTA	4110
Oy	4009	attccctataaagcccgacctcctctcaactacgtgtgtaagaagcccatacaagtlcttc	4068
Db	4111	ATGCAATTTAAACATATATCTCTCAATGATATCATTTATTTGAGAGATTCATTTAAATTTATG	4170
Oy	4069	aataaagaagaaggatlaaatctgtctagtctaacataatacaagaactccatgcac	4128
Db	4171	AATTCAGACAAAAAACAACCTTTAAAGTTTACAAATATATATAAAGAAATCACTAGTAAA	4230
Oy	4129	accgaatcaaatcgcctaaatgtatgtgtgggtatcaacaagatccttgacgaataatcac	4188
Db	4231	AATGATATTAATTTTGCACAGGAAGGTATAGTTATTTAGAAAAGGTTTTAGCGAAATAT	4290
Oy	4189	aagttcgaccttgactctataaagaatatacaagataag-----	4230
Db	4291	AAGGATGATTTAGAATCAATTTAAAAAAGTATCAAGAAAGAAAGAGTTCCTCATCAACA	4350
Oy	4231	-----caagcgagaatgaaaaatat	4251
Db	4351	CCACCAACAACCTCGTGCACGACGAAAAACGAGCAACAAAGAAAGAAAGTAAAGTTC	4410
Oy	4252	ctgcacctctcgtgaataacacgaagccctgtcaagaacgtgaaagcaaaatgcacac	4311
Db	4411	CTTCCATTTTAAACAACATTTGAGACCTTATACAAATACTTGTTATATTAATTTAGCAT	4470
Oy	4312	ttcglaatcaactcgtgagagccaagttcctaactatacttacgagaagagcaatgtgga	4371
Db	4471	TACTTAATTAACCTTAAAGGCAAGGATTTAAGATTTGAATGTTGAAGAAAGATGAACACAT	4530
Oy	4372	gttaaaatcaagggcgtgaactaccctcaaaagaatccaagaagaagctggcagattcaag	4431
Db	4531	GTTTAAATTAACCTTAACCTTAGTGATTTTAAAGCAATTTGATGACAAATTAATCTTTTAAA	4590
Oy	4432	aaaaatacaaatctcgtcggaaatgtgacagcctgtctacgattataaacaacaacaatc	4491

Db	4591	AACACTAACGACTTGGAGSACATTTAAAAATTGATTAATAATGATGATACGAAAAAAGATTAG	4650
Qy	4492	ctgaccaagttctctgccacgtgcatggtgtctcgaaaacctcgccaaaacagtgctgagc	4551
Db	4651	CTTTGGCAATTAATCTTAGTAGACAGGATTAG--TTCAAAAATTTTCTTAATCAATAATATATCA	4707
Qy	4552	aatcgctcgacgagcaacctgtaggcatgctggaacatctccacgaccacatlgcgtgaag	4611
Db	4708	AAATTATTTGAAGGAAAAATCCACAGATATGTTAAACATTTTCACACACCAATGCGGTAAA	4767
Qy	4612	aaacagtgccccacgaataagcgctgtgttcagcatctlggacgagcgcggaagtgaag	4671
Db	4768	AAACATGTCGCGAANAATTCGTGATGTTTCACACATTTAGATGAAGAAAGACMAATGTAAA	4827
Qy	4672	tgctcccgaaactacaacaagaagaagataagtgctgtagaagaccacaccttaccctgc	4731
Db	4828	TGTTTATTTAAATACCAACACAGAGGTATTAATGTGTGTTGAAATTCCAATCTACTTGT	4887
Qy	4732	aatgaaaacaatlgcgcgggtgtgacgcgcgaatgctaaatgacccgaggaagacagcgctct	4791
Db	4888	AACGAAATTAATGAGGAGTGTGATGCAGATGCCCAATGTACGGAAGAAATTTCCAGTAGC	4947
Qy	4792	aacggaagaataacactgtagagtgctactaagcccgactcctatccactctcttgagcgg	4851
Db	4948	AACGGAAGAAAAATACACATGTAATGATCTAAACCTGATCTTATTCACATTTTCGATGCT	5007
Qy	4852	atttttgcctcagacttaattctctgggcattctcctctcgtacatccatctgcatgc	4911
Db	5008	ATTTCCTGGAGTTCCTGTAACCTTTGAGGAATATCATTTCTTATTAATCACTCAATGTAA	5067
Qy	4912	ctgtacagctcatcactaa	4929
Db	5068	TTATACAGTTTCATTTAA	5085

```

RESULT 12
LOCUS PP190G1 5392 bp DNA INV 22-JAN-1999
DEFINITION Plasmodium falciparum p190 gene for surface antigen precursor.
ACCESSION X05624
VERSION X05624.1 GI:4186080
KEYWORDS antigen; glycoprotein; p190 gene; signal peptide; surface antigen.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS 1 (bases 1 to 5392)
Tanabe,K., Mackay,M., Goman,M. and Scaife,J.G.
TITLE Allelic dimorphism in a surface antigen gene of the malaria
parasite Plasmodium falciparum
J Mol. Biol. 195 (2), 273-287 (1987)
JOURNAL 88011243
MEDLINE 2 (bases 1 to 5392)
REFERENCE Tanabe,K.
AUTHORS Direct Submission
TITLE Submitted (21-JAN-1999) K. Tanabe, Osaka Institute of Technology,
Biology, 5-16-1 Ohmura, Asahi-Ku, Osaka, 535-8585, JAPAN
REMARK Sequence revised by author
COMMENT On Jan 26, 1999 this sequence version replaced gi:9926.
FEATURES
source
location/Qualifiers
1..5392
/organism="Plasmodium falciparum"
/strain="MAD20"
/db_xref="taxon:5833"
/clone_lib="(lambda)NM1149 genomic"
136..192
136..5241
/note="p190 precursor"
/codon_start=1
/protein_id="CAA29112.1"
/db_xref="GI:4186081"
/translation="MKITFFDCSLFPIITQCVTHTSYQELVKKLEALDEVALGTYS
LFQKPELVNGTSGTAVATSTPGSSGVYSGVASVASGSGSVASGSGSNGNS
LRFPKSNSSDNGTKRYADLKHRYONTYFTIKELKYLFLDITLHMLILSKNVGGFKYK

```


[illegible]

Query Match	26.1%	Score 1290.4	DB 96	Length 5392
Best Local Similarity	56.4%	Pred. No. 9.4e-265		
Matches 2904	Conservative	0	Mismatches 1966	Indels 246
				Gaps 17
Oy	10	atgaataacattcttcctcctcgtcattcgtttttatcatcaaatcctcagtcgctg	69	
Db	136	atgaagaatcatatttctttttatgattcaattctttttttttattatmaaatcacaaagtgtga	195	

QY	70	accacgaatccctacgaagcvtglttaagaaacctggaacgtttggaagatgcgcgtcctt	129
Db	196	ACACATGGAAGTGTTCAGAAAGACTGTCAAAAAAATCTGAAAGCTTTAGACATACAGTATTTG	255
QY	130	accggaataacgcgtgtcccggaagagaatgtgtcgaataaagagacgcgtgtggcgcg	189
Db	256	ACAGGTTATAGTTATTTTCAAAAGAAAAATGGTATTTAAATGAAGCAACAGTGGACAA	315
QY	190	gccgcttaaacacgaacacccgcggtctcgaagggctcgtggtcgaagctggcctcggtg	249
Db	316	GCGTTTACAACTAGTACACCTGGTTCGAAGGCTTCAAGTACTTACAGTGGTTCACTGGCT	375
QY	250	tctgtggcctctggggggttcgcgtccgcgcggagcgtgtgcataagtggtcgaatg	309
Db	376	TCAGTTGCTTCAGTTGGTTC-----AGGTGGTTCAAGTGGCTCAGTT	417
QY	310	gcaagcgcggttcccggaagaaagtcgagagaccatccatctcgaaacgtcgaatcc	369
Db	418	GCTTCAGGTGGTTCAAGTAATTCAGACGTAACAATCTTCAGATATTTCAAGTAATCA	477
QY	370	gacgcgaagctccacgcgcgacctcaagcacgcgagtgaagaacatctccctacatac	429
Db	478	ATACTAAACCTTACGGCGATTTAAACATATAGAGTTCAAAATTTACTGTTTCACTTTAA	537
QY	430	gagctgaagatcccaacagttgttcgcacctatacatatgtcgacactgtgtgatac	489
Db	538	GAACTCAAAATTCGCCGAACCTTTGATTTAACCAATCATATGTTGACTTTATCTTAAAT	597
QY	490	atctagtcctcaaatctcgaattgagcgtttaagagagatacgaatactcctgtacaag	549
Db	598	GTTGATGGTTTCAAAATTTTATATGATGGAATGAAGAAATTTATGAAATTTATTTAA	657
QY	550	ttagattctctactgcgactgtcgaagggccaacactgaatgacgtttgcgcgaatgact	609
Db	658	TTAAACTTTTATTTATGATTTTATTAAGAGCCAAATTTAAATGATGATGCTCCATATGTTAT	717
QY	610	tgtaaatccatccaattcgaatltgaagatcagaagccacgagtttgagcgtaltgaagaatg	669
Db	718	TGTCAAAATACCTTTCAATCTTTAAATTTCTGCAAAATGAATTTAGACCTACTTTAAAAAT	777
QY	670	gctctcggaatctgcgaagcctctcgacaacataagagcaatgttggagaagtggaagat	729
Db	778	GGTTCGGATTTAGAAAAACCTTATGACAAATTTTAAAGATATCTAGCAAAAAATGGAAGAT	837
QY	730	tatatataaagaataagaaagccatcgaaacatttaacgagctgtgaagaatcccaa	789
Db	838	TACATTTAAAAAATTTAAACACCATATGCAAAATATTAATGAATTAATTGAAGCAAGTTAG	897
QY	790	aagacacatagacaaaataaagaatctgcaaccaagagagaaagaagaagatltgcag	849
Db	898	AAAACAATTTGATCAAAATTAAGAAATGCGATATATGAAGAAAGAAAAAATTTATACCA	957
QY	850	gcccagctcgccgctgtccatctataaacaacagcttgaagaagcccatcactcctcgc	909
Db	958	GCTCAATATTAATCTTTTATTTATTAACAATTAACAATTAACAAGACACATATATTTAATTAAGC	1017
QY	910	gtactcgggaagcgcataagaaacccctcaagaagaatgtgaataatcaagaacgtgtcgac	969
Db	1018	GTTTATAGAAAAACGTATTTGACACTTTAAAAAAAATGAAAAACATAAGAAATTTACTTGA	1077
QY	970	aagaltatgaatlaagaatcctccgcga--gccaactctggaacacccctacaagc	1026
Db	1078	GATATAGATTAATAATTAACAGATGCGGAATAACCCACATACGTGAAGTAAACCAATCT	1137
QY	1027	ctgcctcgagcaagaacagaatagagag--caagagaaagagatccaagagctgc	1083
Db	1138	CTCCCTGTGAATTAAGAAAAAGAAAGTCGAGGACACGAAGAAAAAATAAACAATTTCC	1197
QY	1084	aaaacactaagatccaatagatctctcttaccgatccctctgagcttgagtgatcac	1143
Db	1198	AAAACATATTTAAATTTAAACATTTGATGATTTATTTACTGATCCACTGCAATTTAGATATTTAT	1257
QY	1144	ctgagaggaagaaataagaatatagaacatctccgcaaatgcga-----gacaag	1194

Q7	1021	ccgcgcggcgaacgaacgaagaaatagagagag	caaggaagaagatcagaagatgcgc	1100
Db	1138	CTCCCTGGGAATTAAGAAAAAGAAAGTCGACGAGCACGAAGAAAAAATAAAGAAATTTCC	1197	
QY	1084	aaacacttaagttcaacatagatctctctttacgatcccttgagtggaagrac	1143	
Db	1139	AAACTATTTAAATTTTAACTGATGATTTTATTTACTGATCCACTGGAATTAGAATATTAT	1251	
QY	1144	ttgagagagaagaaataagaatalatagaacatctccgacaaagtcga	1194	

Db	1258	TTAAGAGAAAAAATTAATAAAAGTTGATGTACCTTAATTCACAAAGATCTTACGAATCT	1317
Qy	1195	gaatcaaccgagacctaaatgaatacccaatggtgtagacccctcgtctatataacgat	1254
Db	1318	GTTCCAAATACCAAAAGTTCCGTATATCCAAATAGTATGTATATCCCTTTACCACTACAT	1377
Qy	1235	atcaacaacgctc-----aacgagctcaatagcttcgigtgaactgatacccc	1305
Db	1378	ATTGATATATCATTTAGCTGCGAGATTAATGATATAAAATTCATATGCGATTTAATGAATCT	1437
Qy	1306	ctcgattatacgaagaagaccctctaagaatctacaacgacaaatgagagaaagaagttc	1385
Db	1438	GATCTCTAAGAAAAATTTAATGAAAAATTTATTACGATTAATPACGAAGAAAAATATTTC	1497
Qy	1366	atcaacgaagatcaagggagaagatccaatctgaagaagaagaatctgaagtgacaagaa	1425
Db	1498	ATTATATATACATTTAAAAACAATTTGATTTGAGAGAAAAAACATTATACACAAAAAGAA	1557
Qy	1426	agttacgaagacccgacgaanaagtlciaaacgatalactaaaglatgaanaagctgtc	1485
Db	1558	CAAAATATAAAATTTACTTGAAGATTTATGAAAAGTCAAAAAGAGATTAAGAAGATTA	1617
Qy	1486	aacgagatctatgtatccaaatccaataacacatcgaccctgacccaactcggagaatg	1545
Db	1618	GAAAAATTTTATGTAAGAAATTTAATTAATTAATTTGACAAAAGATGCTGTAGTAAATA	1677
Qy	1546	atgggaagaacgctcctctcaagtgggagaacatgacacatatacctcttgatcc	1605
Db	1678	TTCACTGCGACAGATTTACATTTAATGTTGAAAAACAAGATTAATTAATTAATTTTCACTC	1737
Qy	1606	tatgagaattctaaacataactcttgagaagctccacaagcctctaatgatalatgaggac	1665
Db	1738	TCTATATATTTCTGATTAATTAATGTTCAAAATTTAAAAAAGCTCTTTCATATCTGAACAT	1797
Qy	1666	tattctctgoggaaactgtgtgaggaaagaactaaglatcaagaatccataagt	1735
Db	1798	TATCTTTTAAACAAAAGCAATTTCTGAAAAAGATTTTAACTATTATTAACCTTGAAACT	1857
Qy	1726	aagctcgaaacgagatcgagagcgtgtgtgagaacatlaagaagatggaagaacagtg	1785
Db	1858	GGCCTCGAAGCTGATNTATAAAAAATTTACAGAGAAATTAAGATGTGAAAAACAAATT	1917
Qy	1786	cttgagaagaagattcaaaaaagacgaanaataaacagatctgaagaatccctgagctcc	1845
Db	1918	CTAGAAAAAATTTTAAAGACTACACATTCGCAATG---CTTCCTTGAAGTATCT	1974
Qy	1846	gatatgttaagtcgaagtgcagaagtggtcctctcaatgaacaagatgtgaactcaag	1905
Db	1975	GATATTGTAAAAATTACAAAGTACAAAAAGTTTTTATTAAATAAAAAATGGAAGACTTAAGA	2034
Qy	1906	aagctcaactcatctcgagaagacggtggatlaaaacataatatactatgctcgaaatg	1965
Db	2035	AAGTATAGATTTATTTTAAAAAATGCAACACTTAAAGATGATTTATCTGTACCAATATT	2094
Qy	1966	tataagcagagataaagcagagacacatacctcatctgatactcaagaagaagatagac	2025
Db	2095	TATTAACACAAATTAATTAACCAAGACCATTTTATTATTTGATTTAAAAAAGAGTATGAT	2154
Qy	2026	aaactgaaagtgtcatgcccagaagtcgagagcctgataacgaagagaagaagaacatt	2085
Db	2155	AAATTTAAAGATTTTATATCCAAAGTAAAAAGACATGTTAAAGAAACAAAGCTGTCTTA	2214
Qy	2086	aaaactgaaggaagctc-----agataactcggagcctccacaagaaggagataaac	2139
Db	2215	TCAAGTATTACACCACTTTTATGTTGACGAAGGAAACAACTGAAGATGTGGGCTCACTCC	2274
Qy	2140	ggaacagctaccacaagcccggaacaacagccggtctagactcogaaggcgtatagcgt	2199
Db	2275	ACACACACATTATCCCAATAGCAGAAACAGAACTAACGAAGAAACAGAACTAACGAA	2334
Qy	2200	caagctaaagacagaagcagaagcagagacagcctccagtgccagtgcccgcttc---	2255

D	b	2335	GAAACGTAAGCAGCACACACACAGCGTATACAATTAACTTTACCACCAAAAAGAAATACACAA	2394
Oy	2256	-----agagcgctaaagcctaagtgcttacaccaacgcgcttcgtgatatacaagaacggag	2310	
D	b	2395	CCAAAGAAGATAAAGTGTTGTGAANAATTTCATATGACACATAGAGTAATGACAAATTACAA	2454
Oy	2311	aatgtcagcaaacctcgcacttaccttggaagcttatagattccctgaatacacttaac	2370	
D	b	2455	GCCTTGACAAAAAACGGTTTTCTTAAGGAAATTTGATGAAATTTTAACTAAATCATATATA	2514
Oy	2371	tgccacaatatatcccgcgtctccaagacatbaaaggagaagatlcttaaaagctac	2430	
D	b	2515	TGTCTAATATATATTTTAGTATCAACTCTAGTATGACCAAAAATTTATTAGAGSTATAT	2574
Oy	2431	aagataaacaagaagagagaglaaacctgtccctcttgtatccactcgcgaccctcgtctc	2490	
D	b	2575	AATCTTACTCCAGAAAGAAAGAAAGATTAATAATCATGTGATGCCATTTAGATTTATTATT	2634
Oy	2491	aatatccagaaacaacatccccgtatagtatctatgtlcatagctcaacaactctc	2550	
D	b	2635	AATATTCAAATTAACATACCTCGTATGTATTCATTTATGATAGATGACAAATGATTTA	2694
Oy	2551	tctcaactgltcacvgagagatatatgaaagagatgctgtcgaacctatataactcaa	2610	
D	b	2695	CAACATCTCTTTTTTGAAATTAATTAACAAAAGAAATATTTATTATTACATTAACATAAA	2754
Oy	2611	gacaaacgaagattaagaaccttcggag-----	2641	
D	b	2755	GAGGAAATCATCATCAAAAAATTTTAGAGAGCAAAAAAATAATACGTGAAACATCATCT	2814
Oy	2642	-----aagctaaagaagtlctcaacctcgttaaactctcttc	2681	
D	b	2815	ACATCAGTCTCTGGAATTAACACCGTAATACGTCTCATCCGAACTCACAGTAATTC	2874
Oy	2682	cagctccaatgaaacctcgtctcgaacctcaagaaa-----	2720	
D	b	2875	CAAAACCAACATTAATAATGGATCCCTTACCAATACCCAATAATGCTAGCTGTATCATCT	2934
Oy	2721	-----gcccgagagcgctcctaagcaaacctctcaactcgacaaccttaataactca	2755	
D	b	2935	GGTCTGCTGTAGTTGAAGAAAGTCATGATCCCTTAACGATTTGTCTATTAGTAGCAT	2994
Oy	2776	ctgaaactglttgaaacatccctcgtctcgcgcaagaataaagaacat---ctaccaaga	2832	
D	b	2995	TTGAAGATATTGTTAGTCTCTTAATAATCTGGAATAAATAAGTAAGTACCTAATCATTTA	3054
Oy	2833	cttatctgacagaalocgctccggaacctctacagaaagatctcyaaagacagcgacaa	2882	
D	b	3055	ACCATTCTTACMACAGAGATGAAAAATTTTATGACAAATTTTAAAAAATATATCATACC	3114
Oy	2893	tctcatacggagcttcacttaacttcgtbaaactcaaaagcgatgatatacaactcttc	2952	
D	b	3115	TATTTTAATGATGATATCAACAATTCGTAAATCTTAATTTCAAAAGTAAATTACAGTTTG	3174
Oy	2953	aacgatgatctaaacglaagaactcgyaagagacatcaataaagctgaagaagacactg	3012	
D	b	3175	ACC---GAAACACAAAAAATAATGCAATTAATAATGATGAATAAAAAATTTAAAGATACCTTA	3231
Oy	3013	caactgagctcgcactcgtacaaagatacaaacctgaaactcgyaagagactcctcgacaag	3072	
D	b	3232	CAGTTATCATTTGATTTTATTAATAATTAATTAATTTAAATTTGAAATGATTTATTAAATAG	3291
Oy	3073	aagaaagaaacggcgcaagtblaagatcgagatcaangangltgaactcgtctcaagagcag	3132	
D	b	3292	AAAAAACAATTTGGCCACAGACAAATATGCAAAATTTAAAAAATTACTTTTAAAGAACAA	3351
Oy	3133	cttgaagcaaacctcaactcagaaacaatcogaaacagatctcgaagactcctcagtg	3192	
D	b	3352	TTAGATCAAAATTTGATTTACTTAAATTAACCCACATTAATGATTTACAAAACCTTTTCGT	3411
Oy	3193	tctctcaacaagaagaagcccgagatcgccgagacagagaacactctcgyaagacac	3252	
D	b	3412	TTCTTTTAAACAAAAAAGAAAGCTGGAATTAAGCAAATCGAAATCGAAACACTTTTGGAAACCA	3471

QY	1366	atcaacgaatc	caagggagagatc	caaaattg	gaagaaagaaat	ttgagatg	caagaaga	14235															
Db	1483	ATTATATA	CATTAAAAA	CAAAATTTG	ATTAGAA	GAAGAAAAA	ACATATAC	CAACAAAA	GA	15422													
QY	1426	agttacgaag	ccgcag	caaaag	lctcaaa	gatactca	ctaagag	latgaaga	agctg	14855													
Db	1543	CAAAATTA	AAAAATTACT	TGAAAGAT	TTGAAA	AAATCTCA	AAAAAG	GAATATG	CAAAATTTACT	16022													
QY	1486	aacgagat	ctatg	atccaat	lccaac	taacat	cgact	cccaact	ctgagaa	tg	15455												
Db	1603	GA AAAATTTT	TGAAATG	CAAAATTTT	TAATTA	TAATTTTG	CAAAAG	ATGCTG	TATTAATA	TA	16622												
QY	1546	atgggaaaa	cggta	ctctc	tcaaa	gtgg	gagaa	ctgac	accata	tata	ccctg	16055											
Db	1663	TTCA	GTGCA	AGATAT	TACAT	TATTAAT	GTGAAAA	CAAAAG	TATTAATA	TAATTTT	CAATCC	17222											
QY	1606	latgaga	atctc	taagc	taact	cttg	gaagct	ccacaag	ctctta	g	lata	tg	agag	16655									
Db	1723	TC	TATATATTTCT	TGATAT	TATATTTG	TCAAAAATTT	AAAAAG	CGCTTTT	CAATCTT	GTGAAAT	17822												
QY	1666	latctc	ctcg	gaac	atctg	tctg	ggagaa	gaacta	agat	lata	caag	atc	tca	taag	17255								
Db	1783	TATCTTT	TAA	GA AAAAG	AATTTCT	CGAAAAAG	ATTTTA	ATCAT	TATTTAT	ATCTTT	GA AAACT	18422											
QY	1726	aagat	cga	aaa	cggag	atc	cgag	cgctt	gtt	tg	gaac	atla	gaag	atg	ga	ag	ca	gt	17855				
Db	1843	GGCCT	CGAAGCT	GTAT	TATTA	AAAAA	ATTAC	GA	GA	GA	ATTA	AGAA	TATTA	AGAT	GTAG	TA	GA	AAAA	CAAA	ATT	19022		
QY	1786	ttg	gaaga	gaat	lata	caaa	agc	gaa	ataa	cc	gag	atg	gaag	atc	ctg	ag	ctc	18455					
Db	1903	CTA	GA AAAA	TTTTTT	AAAA	GGCA	CTA	TAC	ACAT	CTG	CAAAA	GTG	---CTTCC	T	TG	AA	AT	TATCT	19595				
QY	1846	gata	tctg	tta	agtc	ccag	cgc	gag	cgag	tg	ctcc	tca	ga	aa	ga	at	tg	at	ga	act	ga	19055	
Db	1960	GAT	TATG	TAAAA	ATTACA	GATAC	CAAAA	AGTTT	TATTA	TATTA	AAAAA	ATG	AG	AC	CTT	T	A	GA	20195				
QY	1906	aagac	tae	act	cat	ctc	tg	tga	aag	cgt	tg	ga	gtlta	aa	a	ca	ta	tata	cat	gtg	cc	ga	19655
Db	2020	AAG	AT	AGAT	TATTTATTTT	AAAAA	ATG	CA	CAAC	CTA	AAAG	TATG	ATTT	CAT	GTAC	T	AC	CAAA	TATTT	20795			
QY	1966	tata	g	cag	gag	ga	atla	g	cag	ga	ac	tact	act	cat	ctg	tact	ca	ga	ag	ag	at	ga	20255
Db	2080	TAT	AA	CC	CA	CAAA	TAA	CA	GA	CA	CAT	TATTTATTT	ATG	TTT	AAAAA	AG	AA	CT	AT	213395			
QY	2026	aa	act	g	aag	atg	tct	cat	g	c	ca	aag	c	cg	a	g	c	t	ca	a	a	g	20855
Db	2140	AA	ATTT	AAAA	CA	ATTTT	TAT	CA	AA	GA	T	TA	A	A	A	A	A	A	A	A	A	21995	
QY	2086	aaa	act	g	aag	a	g	a	c	a	g	a	c	a	g	t	-----	aga	t	a	a	c	213395
Db	2200	TCA	AG	TAT	TAC	ACA	CACTTT	AG	TTT	GC	CA	GA	CG	CA	AG	CA	CA	CA	CA	CA	CA	223395	
QY	2140	gga	c	a	g	a	g	c	t	a	c	a	c	a	a	a	a	a	a	a	a	a	21995
Db	2260	ACA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	23195	
QY	2200	ca	a	g	c	t	a	a	g	c	a	a	a	a	a	a	a	a	a	a	a	22555	
Db	2320	GA	AA	CA	CT	AG	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	23795		
QY	2256	-----	aga	g	c	t	a	a	a	a	a	a	a	a	a	a	a	a	a	a	23105		
Db	2380	CCA	AA	GA	AG	TAA	AG	TAA															

Db	2560	AATCTTACTCCAGAGAAAGAAAATGAAATTAAAAATCATGTGATCCATTGATTTATTTATTTT	2619
Qy	2491	aatctcaagaacaacatccctcgtatgattctcagtcgtcgaatgcctcaacatctctc	2550
Db	2620	AATATTCAAATTAACATACCTGCTGATGTATTCCTTAATGATAGATGACATTCGATTTA	2679
Qy	2551	tctcaactgtctcagtgagatatactgaaagagatgctgcgaacctgtataaactcaaa	2610
Db	2680	CAACATCTCTTTTGTGAATTAATTCAAAGGAATGATTTATTTATTTACATTAACCTAAA	2739
Qy	2611	gacaaagcaaatgaatgaacctctgagag-----	2641
Db	2740	GAGGAATTCACATCAAAAATTTATTAGAGAGACAAAACAAATTAATGAAATCATCT	2799
Qy	2642	-----aagctagaagagctccacctgtctaaacctctctc	2681
Db	2800	ACATCCAGTCTTGGAAATTAACACCGTAATTAATGCTCAATCCGACATCTCACGTAAATCC	2859
Qy	2682	cagctccatctgaacacactgtctctcaacctcaagaacaa-----	2720
Db	2860	CAAAACCAACATCAAAATGATCTCTACCAATACCCAAATAGGTGTAGCTGTATCATCT	2919
Qy	2721	-----gcccgaaagtgagcgtctaaagacgacacctctcaactgcgaaccttaataactca	2775
Db	2920	GGTCCTGCTGTAGTTGAAGAAAGTCATGATCCCTTAACGTATTTGCTATTAGTAACGAT	2979
Qy	2776	ctgaaactglttgagacaatccctgtctctcgcgcaagaataaagaacat---ctaccagaa	2832
Db	2980	TTGAAAGGTATTTGTACTCTTTAAATCTTTGGAAATTAACCTAAAGTCTCAATTCATTA	3039
Qy	2833	ctaatgtgacgaatctgctccggaacctctcaagaaagatactgaagacagcgacaa	2892
Db	3040	ACCATTTCTACACAGAGATGGAAAAATTTTAAGAGAAATTTTAAAAATTAATGATACC	3099
Qy	2893	tctataacgagagcttcaactaaactctgtgaaactctaaagccgagatatacaactctct	2952
Db	3100	TATTTTATGATGATATCAAAACATTCGTAAATCTAATTTCAAAAGTAATTAACAGTTTG	3159
Qy	2953	aagcatgaatctaaagctaaagacgagagagacataaagaactgaaagagagacgtg	3012
Db	3160	ACC---GAACACAAAAAAATGCAATTAATGATGAATTTAAAAATTAAGATATCTTTA	3216
Qy	3013	caactgagcttcgcacgctgtacaaagaatacaactgaactgagagagctctgcgaag	3072
Db	3217	CAGTTATCATTTTGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAG	3276
Qy	3073	aagaagacagtcggaagctaaagctgaaagcagatcaagaagctgactctgcgaagagag	3132
Db	3277	AAAAAAGAACTTGCCCAAGCAAAATGCAAAATTAATAAACTTACTTTATTAAGAACAA	3336
Qy	3133	cttgaagcaaatcaactcaactgaacaaactcgaaacacgtactgagagaaactctcagtg	3192
Db	3337	TTAGATCAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3396
Qy	3193	tctctcaagaagaagagagacgagatcccgagacagagaaacactctgagaaacac	3252
Db	3397	TTCTTTAACAATAAAAGAAAGCTGAAATTAACGAACTGAAATAACATTTGAAAAACACA	3456
Qy	3253	aagatctctctcaaacactcaaaagccctcgtcaagctatataagagagagctctctct	3312
Db	3457	AAAAATATTAATTAAGAAACATTTAAAGACCTTGTTAAATTAATTAATTAATTAATTAAT	3516
Qy	3313	ctgaagactctctcgcgagagagatccgagacgagagataactacgccaagcctcgagac	3372
Db	3517	TTAAAAACTTTAAGTGAATGATCAATTTCAAAACGAAGATTAATTAAGCAAAATTTGAAAA	3576
Qy	3373	tctcaagctccgtcttaagcctcgaaagcgaagctgaaagacaaactgaaactctgagaaag	3432
Db	3577	TTTAGAGCATTAAGTAAAAATAGATGGAACAACTCATGATTAATTTCAATTTTAGGAAGAAA	3636
Qy	3433	aagctcagctactcctctagagactgactcaacctgactcgcgagcctcaagaaagctatt	3492

Db	3637	AAATTAATCTTTCTATCAAGTGGATTACATCAATTTAATTACTGAATTAATAAGAACTAATA	3696
Oy	3493	aagaaacaagaactacacccgcgaatagcccaagcgcgaataataacagaagtgataacgca	3552
Db	3697	AAAAATAAAAATTATACAGGTAAATTTCTCCAAAGTGAAAAATATATAGAAAGTTACACAGCT	3756
Oy	3553	ctggaaactcttaacaagaagttccctgcctggaaggaagaagatgctgcacatgtgtgtgc----	3608
Db	3757	TTAAATATCTTACGAAATTTTCTCCGAAAGCAAAAGTTACAAACAGTTGTAACTCCACCT	3816
Oy	3609	-----tgatctgtctccgacacactlgaagcagtlca	3641
Db	3817	CAACGAGATGTAACATCCATCCATTATCTGTAAAGGTTAAAGTATAGTTACAGATCCACA	3876
Oy	3642	actcaagaa-----gcctgcatactcaatgtctgcagccgagtlccaatacaattacc	3693
Db	3877	AAAGAGAAACACAAATATCCACACTTCAGGCTCTTTATTATACGAAATTAACACAACTAGTA	3936
Oy	3694	acatctcaagaacgcgcgcgcgtlgaagtgatgaagtcataatgtgtcctactctctggcag	3755
Db	3937	CAATTACAAATTAATGACGAGAAAGATGATCTCTTAGTGTATTTACCCATTTTGTGGACAA	3996
Oy	3754	agcgaagaggactcagaatgacccctgcgcgaagtgatgcaccgcgtlgaagcgtlcaactcttc	3813
Db	3997	TCCGAGATATATGACGAATATTATAGATCAAGTAGTAGTAACGGAAAGC---AATATCTGTC	4055
Oy	3814	gtgatatacaactctgtcccaaaatcgagacgaatacgaagtgctctatactgaacct	3873
Db	4054	ACAAATGATAATATCTCTCAGGATTTGAAAAATGAATATGATATATATTTAAACCT	4113
Oy	3874	ctggcagagcgctatagtgctctcaagaagaagcctlgagaataactgtatgaacctcaat	3933
Db	4114	TTACTGTGAGATATATACACTTTAAAAAACCAATTGAAAAAACCTTTATCAATTATAT	4173
Oy	3934	gtcaacgtlgaagacatctctgaaacagccgcttaataagaagaaatctcaagaacgctc	3993
Db	4174	TTAATTTGAAACGATATCTTAAATATACGCTTAAGAAACGAAATATTTCTTAGATGA	4233
Oy	3994	ctlgagagcgactgtatccctataaagacgtgaacctcctctaactcgtltgaagac	4054
Db	4234	TTTAAATCTGATTTTAAACGAAATTTAAACATATATCTCTCAAAATGAATATCTTTGAAAT	4293
Oy	4054	ccatacagttccccaataaaggaagaggaataaattctgtctagtatacaactatc	4113
Db	4294	TCATTTAATATTATTGAATTCAGAACAAAAAACACTTTTAAAAAGTTACAAATATATA	4355
Oy	4114	aaggactccatcgacacacgaatatacaattctgcctaaatgatgtcgtgggattatacaagtc	4173
Db	4354	AAAGATACAGTAGAAATATGATTTAAATTTCGACAGAGAGATATAGTTATTATGAAGAAG	4413
Oy	4174	ctlgacgaanaatacaagtlctgacactgtactatataaagaalatacaacgaataag---	4230
Db	4414	GTTTTCGCAAAATTAAGATGATTTATAGATCAATTTAAAAAGTTATCAAAAGAAAGAAG	4473
Oy	4231	-----caaa	4233
Db	4474	GAGAGATTTCCATCATCACCACCAACACACTCGCTCACCGCAAAAAACAGACGAAACAA	4533
Oy	4234	ggcgaagaatgaanaatactgcgcctctcgttaataacatgaacacctgtacaagaacgtg	4293
Db	4534	AAGAAGGAAGTAGTACTCTTCCATTTTAAACAAACATTGAGACCTTATACATTAACCTTA	4593
Oy	4294	aacgaacaaatcgacactcttcgttaatctcaactcctlgagccgaagtgctcctaactatactac	4353
Db	4594	GTTAATTAATAATGACGATCTTAAATTAACCTTAAGACCAAGATTAAGCATGTGTAATGT	4655
Oy	4354	gagagaagcgaatgtggaagttaaaatcaagaagcgtgaactaccctcaaaaacatccaagac	4413
Db	4654	GAAGAAAGATGAAGNCATGTTTAAATAATACATAAACTTAGATTTTAAAGCAATGTATGATAC	4713
Oy	4414	aagtgccagatctcaagaanaaaatacaattctgcgtgaattgcgcagactgtctacgat	4473
Db	4714	AAATATAGTCTTTTAAAAACCTTAACGATTCGAAGCAATTTAAAAAATTGTATATATAT	4773

QY	4474	taaccacacaaatcccccgcaccgaattctgctccacgagcatggtctgcaaaactc	4533
Db	4774	GATACGAAAAAGATATGCTTGCCAAATTAATCTAGTACAGGATTAG--TTCCAAATTTT	4830
QY	4534	gccaaaacagtcgtgagcaatcgtctcgacggaacctgcagggcatgctgaacatccc	4593
Db	4831	CCCTAATTCATTAATATCAAAATTTAATTGGAAGGAAATTCGAAGATATGTTAACTTTCA	4890
QY	4594	cagcacaatgctgtaagaacacagctgccccagaaatagcgtgtttcaaggcatctgac	4653
Db	4891	CAACCCAAATGCGTAAAAAAACAATGTCCAGAAAAATTCGATGTTTCAGACATTTAGAT	4950
QY	4654	gagcgagagatgtagcaagctgtcccgaaactcaaacgaagaagagataagtgctgag	4713
Db	4951	GAAGAAGAAAGATGTAAATGTTTATTAAATTTCCAAACAGAAAGTGATTAATGTTTCA	5010
QY	4714	aaccacaacctaccctgcacatgaaacaaatggcgggtgtgtagccgcatgctaattgacc	4773
Db	5011	AATCCAAATCCTACTTGTTAAGAAATTAATGATGGATGTGATGACGATGCCCATGTAC	5070
QY	4774	gaggaagacagcggtctctaacggaagaaataatcacatgctgagtgtaactaagccgactcc	4833
Db	5071	GAAGAAATTCAGGTACACAGCAAGAAAGAAATCACATGTGAATGACTAAACCTGATCT	5130
QY	4834	tatcacctctgcacggagattttctgcacagctcaattccctggacatccctctg	4893
Db	5131	TATCCTATTTTCATGTGATGTTTCTGTGACAGTCTCTAACTTTTGAAGATATCATTTTA	5190
QY	4894	ctgactcctacgtcatcctgctacagctcatcctaataagatcgatg	4939
Db	5191	TTAATCTCAATGTTAATATATATATACCTTTCATTTAAAAATGTAGG	5236
RESULT 14			
PMESZSA1B			
LOCUS	5312 bp	DNA	INV 10-MAR-2001
DEFINITION	P.falciparum gp190 (MSA1, MSP1, PMMSA) for precursor of major		
ACCESSION	235327 U09209		
VERSION	235327.1 GI:929795		
KEYWORDS	gp190 gene; merozoite surface antigen; MSA1 gene; MSP1 gene; PMMSA gene.		
SOURCE	malaria parasite P. falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE	1. (Phases 1 to 5312)		
AUTHORS	Pan, W., Tolle, R. and Bujard, H.		
TITLE	A direct and rapid sequencing strategy for the Plasmodium falciparum antigen gene gp190/MSA1		
JOURNAL	Mol. Biochem. Parasitol. 73 (1-2), 241-244 (1995)		
MEDLINE	96123395		
REFERENCE	2. (Phases 1 to 5312)		
AUTHORS	Tolle, R., Bujard, H. and Cooper, J. A.		
TITLE	Plasmodium falciparum: variations within the C-terminal region of merozoite surface antigen-1		
JOURNAL	Exp. Parasitol. 81 (1), 47-54 (1995)		
MEDLINE	95354793		
REFERENCE	3. (Phases 1 to 5312)		
AUTHORS	Tolle, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (18-JUL-1994) Ralf Tolle, Abt. Prof. H. Bujard, Zentrum für Molekulare Biologie, (ZMBH), Univ. Heidelberg, Im Neuenheimer Feld 282, Heidelberg, 69120, Germany		
COMMENT	On Aug 5, 1995 this sequence version replaced gi:535249.		
FEATURES	Location/Qualifiers		
source	1. .5312		
	/organism="Plasmodium falciparum"		
	/isolate="NF54"		
	/db_xref="taxon:5833"		
	/chromosome="9"		
	/clone="3D7A"		
	/dev_stage="blood stage"		

OY	1387	atbaaatcttggaagaagaaattggagugacagaagaagttaagaagccgcgcaaa	1446
Dd	1549	ATTGATTTAGTAAAGAAAAAACAATTAACTACACAATAAGAACAAAAATTTACTTGAA	1608
OY	1447	agctcaaacgcatcactaaaglatgaanaagctgtcgacagatactgatcccaa	1506
Dd	1609	GATTATGAAAAGTCAAAAAAGGATTATGAAGAATTACTGGAAAAATTTTATGAAATGAAA	1666
OY	1507	tccaacaatacatcgacctgaccacttcgagaaaaatgaatggaaaacygtactcttac	1566
Dd	1669	TTTAATATTAATTTTTGTGACAAAGATGTCGTGATAAATATTCAGTGCAGATATACATAT	1728
OY	1567	aaagtgsgaagaaacgcaacaccctaataacttgcatactcatagagaattctaagaatat	1622
Dd	1729	AATGTTGAAAAACAAGATATAATAATAAATTTTCAATCCCTAAATATTCCTGTATATATAT	1788
OY	1627	cttggaaagctcaccaaaagctttaagtataatgtagaactatctctcgagacattgt	1686
Dd	1789	GTTCAAAAATTTAAAAAAGCCTCTTTCATATCTTGAGAGATTATTCCTTTAAGAAAAAGAATT	1848
OY	1687	gtvgsgaaagaactaaaglatlaaagaatctcataagtaagatagaatgaacagagatcgag	1748
Dd	1849	TCTGAAAAAGATTTTAAATCATATTATTAATCTTGAAAACTGCCCTCGAACCTGATATMAAA	1908
OY	1747	acgccttgtltggaacatlaagaagatagaagacagatgtlittlgagaagaagatctcaaa	1806
Dd	1909	AAATTTAACGAGAAGAAATTAAGAGTAGGAAACAAAAATCTGAAAAAAAATTTTAAAGGA	1968
OY	1807	gacgnaataaaccagatltggaagaatccgtgagatctcgatatgtttaagtlccaagt	1866
Dd	1969	CTAACACATTGACGAATG---GTTCCTTGAGAGTATCTGATATTTGTAAATTTACAGTA	2025
OY	1867	cagaagtygctcctcatgaaacagatltgaaactcaagaagaactcaactcatlctgag	1926
Dd	2026	CAAAAAGTTTATTAATTTAAAAAATAGAGACTTAAGAAAGATAGAAATTAATTTTAAAA	2085
OY	1927	aacgtgsgatlaaacaataataatcatatgbtgcgaataagttataagcagaagaaataagag	1986
Dd	2086	AATGCACACATMAAAGATAGTATTTCAATGCCAAATATTTTAAACACAAAAATTAACCA	2145
OY	1987	gaaccaatacctcccatacgtactcaagaagaagatagaacaactgaagtgctcagcc	2046
Dd	2146	GAACCATATTTTAAATTTGATTTAAAAAAGAGTATGATTAATTAACAAATTAATTAACA	2205
OY	2047	aaagltcgagagctgfatcaacagagaagaagaacatlaaacctgaaagcagacg----	2102
Dd	2206	AAAGTAAAAAGCATGTTAAGAAGAAACAACGCTGTATACAGATTTACACAACTTTA	2265
OY	2103	--agataactcggagccttccaagaagaagagataaacgsgacaggtcatccaacaagccc	2160
Dd	2266	GWTCACGACAAACGGAAMACACTGGAAGATGGGGGCACTCCACACACACATTAATCCCAATCA	2325
OY	2161	ggacaacaagagccggttcatgctcttcgaaaggagatagagtgcaagctccaagacaagaagag	2220
Dd	2326	GGAGAACACGAAGTAAACAGAAAGAACGANGAANACGAAGNAACAGTAAAGACACAAACA	2385
OY	2221	aagcagagc-----acagccttcagigtgcaggtgccggttccagagaggtcataagctcoa	2271
Dd	2386	ACGGTAACAATAACATTAAACCAACAACAACAACATCACCAACCAAAAAGAAAGTAAAGTTGTT	2445
OY	2272	gtgcctacaacaacagctcctgtggaataacaagaaccggaatgtcagcaaacctgagctac	2331
Dd	2506	CTAAAGAAATTTAGTGAATTTTAACTAAATCATATATATGTCATTAATATATATTTTAATA	2565
OY	2392	tctcaacagcactaagcagagaagatctcttaaacagtaacaagataaacaagaagaagag	2451
Dd	2566	TCAAACCTCTGTATAGGACCAAAAAATTTATTAAGAGCTTATATACTTACTCCAGAAAGAGA	2625
OY	2452	agtaaacgtctccttgtgtgatacactgagacgtgtgtltaatatccagaaacaactcc	2511

[illegible]

Db	3703	GGATTACATC	TTTAATTACTGAA	TTTAAAGAA	AGTAATATA	AAAAATTTAAAT	TATACAGT	3762
Oy	3514	aatagcccaagcgagaa	taatacaga	cgta	gaataacgcac	ctgnaatct	tacaagaatc	3573
Db	3763	AATTCTCCAA	AGTGAAATATAT	TAAGAAAG	TTAAACGA	AGCTTTAAAT	CTTACGAAATTTT	3822
Oy	3574	ctgcctgaagaa	aagatctgcgcac	ctg	g	g	g	3608
Db	3823	CTCCCAACAA	CGMAAAAGTTTAC	ACAGTTGTAA	CTCCAC	CTCAACAGAT	GTAACTCCATCT	3882
Oy	3609	-----Tgaatc	tggtccgcagacac	ctggaagc	gagtc	ttaacat	aaagaag-----cct	3654
Db	3883	CCATTATCTG	TAAAGGTAAGT	GGTAGTTAC	GAGTCCAC	AAAAAGACAA	ACAAATPACCA	3942
Oy	3655	gcactctcaat	gtctcgagcgcag	ctccat	tacat	ttacacat	ctcgaagc	3714
Db	3943	ACTTCAGGCT	CTTTATTAAC	GAAATTAACA	CAAGTAGT	GCATTC	CAAAAATATATGCGAA	4002
Oy	3715	gaagtcga	taagtcatactgtgc	tatctctgcgcga	agcgagga	gaggaactaga	gtac	3774
Db	4003	GAAATGATTC	CTTAGTTGTTAT	ACCACATTTT	TGGAAATCC	GAAGAT	AATACGAAATAT	4062
Oy	3775	ctcgacga	gtgtgtcaccggtga	ggcggtgaac	ccctccg	tgattga	taagaatctgtcc	3834
Db	4063	TTTAATTAAGT	AGTACTGGAAGC	-----AATATCTG	TACATGAT	GTATATCTCTCA		4119
Oy	3835	aaatcga	gaagaatacgaag	gtctctatctg	aaacctctg	gcagcgct	catagatgct	3894
Db	4120	GGATTTGAA	AAATGAAATGTAT	GTATATATTT	TTAAACCTTT	AGCTGAGT	ATATATGAAAGC	4179
Oy	3895	ctcaaga	aaacgcgtcg	gggaataacg	tga	tgaccttaat	gtcaacg	3954
Db	4180	TTTAAAAA	CAAAATGMAAAAC	CAATTTTAC	TTTAAATTTAA	TTTGAACAT	ATCTTTA	4239
Oy	3955	aacgcgcct	tataaagaga	aaatcttaaga	aacgctctg	gagagcagc	ctgtatcc	4014
Db	4240	AATTACGCT	CTTTAAGAA	CAAAAATTTTCT	TGATGTAT	TGCAATCTGAT	TTTAATGCAA	4299
Oy	4015	tataa	gacctg	taactctctat	taactg	tcaagga	cccaatac	4074
Db	4300	TTTTAA	CATATATCTTC	CAATGAA	TAFCATAT	TATGAGAT	CTTAAATATTTGAATCA	4355
Oy	4075	gagaaga	gggataaa	atctctg	ctagtata	caataata	gaagactc	4134
Db	4360	GAA	CAAAAAA	CAACACTTTT	AAAAAGCTT	CAAAATAT	TTAAAGAAATCACTAC	4419
Oy	4135	atcaatt	tccta	atgtgtgc	gggtat	tcaaga	ctc	4194
Db	4420	ATTAA	TTTGAC	AGGAA	GTATAT	AGTTAT	TGAAGAAAGTTT	4479
Oy	4195	gacctg	actctat	ttaaaga	tata	tcaacg	ataag-----	4230
Db	4480	GATT	TAAATC	ATTTAAAA	AGTTAT	TCAAAAGAA	AGAAAGTCCCAT	4533
Oy	4231	-----	caagcg	gaat	tgaa	aatat	ctg	4254
Db	4540	CCAACA	CAACCTCG	CTCACC	AGCAAAA	AGACAGAA	CAAAAAGAAAGTA	4599
Oy	4255	ccctctc	gaataa	atcga	aaacccg	tcaaga	acag	4314
Db	4600	CCATTTT	TAAACA	AACTGTG	AGCTTAT	ACAACTTG	TTTAATTAATTTG	4659
Oy	4315	gtaat	caact	ggagc	gaagc	gtccca	actata	4374
Db	4660	TTAT	ATTAACTTAA	AGCAAGAT	TTTAACGAT	TTGTAAT	GTTGAAAAAGAT	4719
Oy	4375	aaaat	caagagc	tgaa	ctac	ctcaaaa	caatc	4434
Db	4720	AAAAT	TAAC	TAAACTT	AGTGTG	TTTAAAA	AGCAATGATG	4779
Oy	4435	aataca	attctg	tcggaat	tgaagc	ctgtctac	cgatata	4494
Db	4780	CTTTAG	CACTTGA	AGCAATTT	AAAAATTTG	ATATGAT	ATATGCAAAAAA	4839

OY	4435	acaaagttctgcctccactgcacgagtggttcgaaacacctgcgcaaaacagtgctgaagcat	4554
Db	4840	GGCAAAATTAATCTACTACAGGATTAAG--TTCAAAATTTTCTTAATACATAATATATCAAA	4896
OY	4555	ctgcctgcagcgcaacccgcagggcgcgtgtaacatctccccagcaccaatgctgtaagaa	4614
Db	4897	TTAATTGAGGAAAAATTTCCAAGATATGTTAACATTTCACAAACACCATGCGTAAAAAA	4956
OY	4615	caatgcccccaagaaatagcggcgtgttccaagcaltcggacgagcgcgaagatgtcaagtgt	4674
Db	4957	CAATGCTCCAGAAAAATTTCTGGATGTTTCAGACATTTTGATGAAAGGAAGAAATGTAAATGT	5016
OY	4675	ctccctgaactcacaagaagaagatlaaigtggtggaagaccacaacccctacccgcat	4734
Db	5017	TTATTAATAATTCAACACAGAAAGGTGATTAATGTGTGAAAATTCACAAATCTACTGTATAC	5076
OY	4735	gaaacaatgycgggtgtgtagccgcgcatgtcctaaatgcaccgaggaagacagcggctctaac	4794
Db	5077	GAAATATATGTGTGATGTGATGTCAGATGCCATGTACGAGGAAGATTCAAGGTGACGC	5136
OY	4795	ggaaagaaaatcaaatgctgggtgtlactaagcccgactcctalcacactcttgcgaaggatt	4854
Db	5137	AGAAAGAAAATCAACATGTGAAATGTACTAAACCGATTTCTTAATCCACTTTTGCATGATTT	5196
OY	4855	ttttgctccacgtctaaattccctggcgacatccctctcgtgataccatgctgtacctg	4914
Db	5197	TTCTGCAATTCCTTAACCTCTTAAGGAATATCATTCTTTAATATCTCATGTTAATATTA	5256
OY	4915	tacagcttcacatcaatagatcagatg	4939
Db	5257	TACAGTTTCATTTAAAAAATGTATAGG	5281

RESULT	15
LOCUS	PENGAP195
DEFINITION	P.falciparum major merozoite surface antigen (gp195) gene, complete cds.
ACCESSION	M7213.1
VERSION	GI:160315
KEYWORDS	major merozoite surface antigen. P.falciparum (strain Uganda-Palo Alto) DNA.
SOURCE	Plasmodium falciparum Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ORGANISM	1 (bases 1 to 5276)
REFERENCE	Chang,S.P., Kramer,K.J., Yamaga,K.M., Kato,A., Case,S.E. and Siddiqui,W.A. Plasmodium falciparum: Gene structure and hydropathy profile of the major merozoite surface antigen (gp195) of the Uganda-Palo Alto isolate
AUTHORS	Exp. Parasitol. 67, 1-11 (1988) 89005525
JOURNAL MEDLINE FEATURES	
source	Location/Qualifiers 1..5276 /organism="Plasmodium falciparum" /db_xref="taxon:5833" 1..5181 /note="major merozoite surface antigen" /codon_start=1 /protein_id="AA29611.1" /db_xref="gi:160316" /translation="MKIIIFLCSPFLPIINOCVTHESYOELVKLEALPADEVLTGCVLFHREKILNEETITSGASNOGSTSGTSGTSRGSSTSGTSAOCTSGTSNOCTSGTSAOCTSGTSGTSPSSRNLTLPKRSTSSGASPDPADSDSKASTADLKHRYRNYLFTIKELKPELFDLTNHLTLCDNIHGKRYLIDYEETINELLYLNFPDLIRAKLVNDVCAGCOIIPENLKIRANELDVLKLLFYGRPLDNIKDNQKMEDYIKKKTTIANIWLEGSSKKTIDONKNADNEGKKRLVQAQYDLSLYNQOLEAHNLISVEKRIDTKENNEIKELDKINEKNPPANSNGTNPTMLTDKKRIIEHEERIKELAKTIKENIMDSLETDDLELEYLRBKRNKYDYTPRKODPKTSQIQVPYPNPDIYPLPLDTIHNSLAANDKSNYGDLMPNPDTRKEIKENEITIDNKERKFIVNNIKRODLEEKINHREOKRKLELDYRSKSKKLDEYELLKEFYEMKFNENPFKDVAVDKLFSAAYVNVBQRANNKFSFSSNNSYAVNQLKKSLDYLEDYSLEKGISSEKDFPNVYTLTGLEADYIKRLTEIKSSEKSR
CDS	

Db	1741	TCATATCTTGAAAGTTATTTCTTTAAGAAAAGGAAATTTTCGAAAAAGATTTTAATCATATAT	1800
OY	1711	aagaatccataagtaagatcgcgaaaaacgagatcgcgacgtctgtgtgagacataaagaag	1770
Db	1801	TATACTTTGAAAACCTGGCCTCGAAGCGTGATATATAAAAAATTAAACGAAACAAATTAAGAAGT	1866
OY	1771	gatgaagaacagtgctgtgttggaagaagaatacaaaaagcgaataataaacgagatgaag	1830
Db	1861	AGTGAACAAATTTCTAGAAAAAAATTTTAAAGGACTAACACATTCAGCAAAATG---CT	1917
OY	1831	atcccgagaggtcccgatattgttaagtcgaagtcgagaaagtgctcctcctcgaagaag	1890
Db	1918	TCCTTAGAAGATATATGATTTATGTTAAATTTACAAAGTCCAAAAAGTTTTATTAATTTAAAAA	1977
OY	1891	atgatgaactcaagaagaactcaactcatctctgaagaacgttgaggtlaaacataata	1950
Db	1978	ATAGAAAGCTTAAAGAAAGATAGAAATTTATTTTAAAAAATGCAACACTAAACATATGTAAT	2037
OY	1951	catgtgcgaaatagtttaagcagaggaataaagcagaacataactactcctcatgtaac	2010
Db	2038	CATGTACCAAAATTTATTTTAAACCAACAAATAAACCAGAACCATATATTATTTAATGTATTA	2094
OY	2011	aagaagaagatagaagaacaacggaagtgtaacgtcccaaaagtcgaagccttgtaaga	2070
Db	2098	AAAAAGAGATGATGATTAATTTAAAGAAATTTATACCAAAAGTAAACACATGTGTTAA--GA	2155
OY	2071	gagaagaagaactlaaactgaagaagcagtcagataaactccgagcctcccaagaaga	2130
Db	2156	AAGAAACAGCGTCTTTTMCAGATATTACACACTTTAGTTGCGACGACCGAAACACTG	2215
OY	2131	gagataaccggaacaggtctaccaccaaacccggaacaaacaggcgggttcagctctgaagc	2190
Db	2216	AAGAT--GGGGGTCACTCCACACACACATTTATCCCAATCAGAGAAACAAAGATTAACAA	2274
OY	2191	gatagcgtgcgaagtcgaagcaagaagcagaagcagagcagagcctccagtgccagtgccc	2250
Db	2275	GAACAGAGAAAGAAACGAAAGAAACGATGAGACACACAAACAGCGTAAACATTAACATTACCA	2334
OY	2251	gtcccaagaggttaaggtcgaagtcgctacacacacagctccctgtgataaagaagcag	2310
Db	2335	CCAAAAGAGTAAAGTTGTTTGAAGAAATTCATATGAAACATTAAGAGTATATACAAATTACAA	2394
OY	2311	aatgtcagcaaacctggaactcctgtgaagctctatgaggttcctgataatcactacac	2370
Db	2395	GCGTTGACAAAAACAGTTTATCTTAAGAAATTTAGATGAATTTTAACTTAACATATATTA	2454
OY	2371	tgccacaataatccctcgtcctctcacaagcactatgaacagaagaatctctaaacagtaac	2430
Db	2455	TGTCTAATTAATTTATTTATAGTATCAAACTCTAGTATGAGCCAAAAATTTATTAAGAGTATAT	2514
OY	2431	aagaataaccaaggaagagggaggtaaactgtcctctgtgataccactggaacctgtgttc	2490
Db	2515	AATCTTACTCCAGAAAGAAAGAAATGAATTAATTAATCTATGATCATTTAATTAATTAATTT	2574
OY	2491	ataatccgaagaacaactcccgatlatgtaactcgtatgcgataccctcaacaactcttc	2550
Db	2575	AATATTTCAAAATTAACATACCTGCGATGATTAATTAATGATATAGTATGAACAATGATTTA	2634
OY	2551	tctcaactgttcatgtagatatatgataagaagatgagtcgtcgaacctgtataactcaaa	2610
Db	2635	CACATCTCTTTTGTGATATATATCAAAAAGGAATGATTTATTTATTTACATTAACCTAAAA	2694
OY	2611	gacaacgagaagatlaagaacctctcggaggaagcctaagaaggtctc-----	2657
Db	2695	GAGGAAATTCACATCAAAAAATTAATTTATGAGAGCAAAAAAATTAACATTAACATCATCT	2754
OY	2658	-----cactctgttaaacactctcttc	2681
Db	2755	ACATCCACTCTCGAATTAACACCGTAAATACATGCTCAATTCGCAACACTCACAGTAAATTC	2814
OY	2682	cagctcgaatgcaccactgtctcacaacctcaagaacaa-----	2720
Db	2815	CAAAACCAACATCAATATGATATCTCTTACCAATTAACCAAAATATGATGATCTATCATCT	2874

QY	2721	-----gcccgagatgagcgcgtacagacagacccctcactcgacccaaccttaataacta	2775
Db	2875	GGTCCTCGCTAGTATTGGAAGAAAGCATGATGCCCTTAAACAGTATTTGCTATAGTAACGAT	2934
QY	2776	ctgaacctglttggagaacatccgctctcttcggagaataaagaacat---ctaccagaa	2832
Db	2935	TTGAAAGGTATTTGGTTAACTCTTAAATCTTTGGAAATTAACCTAAAGTACCTAAATCCATTA	2994
QY	2833	cttatggacgaatccgctccggaacttcctacgaagatactgaagacagcgacaca	2892
Db	2995	ACCATTTTACACACAGATGGAAAAATTTATGAGACAAATATTTTAAATAATATGATATCC	3054
QY	2893	tctctaaacgaagcttcactaaacttcgltgaatctcaaaacgagatatacaactctct	2952
Db	3055	TATTTATATGATGATATCAAAACAACTTCGTAAATCTAATTTCAAAAGTATATTCACAGTTTG	3114
QY	2955	aacgatgaactlaaacgtaagaagctcggaaagagacatacaataagctcgaaagaagacatg	3012
Db	3115	ACCGGAACACAAAA---AATGCATTTAAAGATGAAGAAATTTAAAAATTAAGATCTTTA	3171
QY	3013	caacgagcttcgcagctcgtaacaacagtaacaactgaactggagactctcgacag	3072
Db	3172	CAGTATATCATTTGATTTATATATTAATATATAAATTAAAAATTGATATGATATTATTAATAG	3231
QY	3073	aagaagacagctcggaagataaagatgcagatcaagaagltgactctgtccaagagcag	3132
Db	3232	AAAAAGAACTTTGGCCAAAGCAAAATGCAATTTAAAAAACTTACTTTATTAAGAACAA	3291
QY	3133	cttgaagcaaaactcaactcactgaacaatccgaacacgctacgcagaaactctcagtg	3192
Db	3292	TTAGATCAAAATTTGAATTTACCTTTAAATACCCACATTAATGTTTACAAAACTTTCTCTT	3351
QY	3193	tctctcaacaagaagaagacccgagatccgcgagacagagaacactctggagaacac	3252
Db	3352	TTCTTTAAACAAAAAAGAGAGCTGGAATATGCGAAGTAAGTAACATTCATTAGAACACAA	3411
QY	3253	aagatctctctcaacactcaaaagcctcgtaaglatataatggcagctctctct	3312
Db	3412	AAAAATATTTGAAACATTTATAAAGGACTTTGTAATATTTATTAATATGTAATCATCATCCA	3471
QY	3313	ctgaagactctctccgagagagcatccagaccgagataactaagccagcctcgagac	3372
Db	3472	TTTAAAACTTTAACTGAAGTATCATTTCAAAACGAAGATTAATTAATGCCATTTAAGAAAA	3531
QY	3373	ttcaagctccgctctgaagctcgaaagcagctgaagagacaactgaactcgtgaagaag	3432
Db	3532	TTTAAAGATTTAACTTAATAATATGATGAGAAACCTCAATGATTAATTTACATTTAGAAAAAAA	3591
QY	3433	aagctcagctacccctctgagcgacgtcacaccctgatccgagatccaagaagatcat	3492
Db	3592	AATATATCTTTCTTATCAAGTGGATTTACATCACTTAATTAATTAATTAATTAAGAACTAATA	3651
QY	3493	aagaaacaagaactcacccggaacatagcccccaagcgagaataatatacagaagtgataacga	3552
Db	3652	AAAAATTAATAATTTATACAGTAATTTCCAAAGGAAAAATTAATGAAGATTAACGAACT	3711
QY	3553	ctggaactctacaagaagcttcgctcgctggaaggaacagatgctgcacgtggtgctc----	3608
Db	3712	TTTAAATCTTACGAAATTTTCTCCGAAAGCAAAAGTTACAAAGTTGATTAATCTCCACT	3771
QY	3609	-----tgaactctgctccggaacacgtgagcagttcca	3664
Db	3772	CACCAAGTATTAACCTCCATTCATTTATCTGTAAAGGTATAGTGTGATTCAGATCCACA	3831
QY	3642	accttaagaag-----cctgcactactcaatgctcgagccgagltccaatacaattacc	3693
Db	3832	AAGAGAGAAACAAATAATACCAACTTCAGGCTTTTATTTAAAGAAATTAACAACTAGTAA	3891
QY	3694	acatctcgaagctcgagagatgaggtcgaatgaacgtcatcaatgtgtcctatcttcggcgag	3753
Db	3892	CAATTACAAATTTATGACGAAGAGATGATTTCTTGTTGTTTACCCATTTTGTGGAA	3951

OY	3754	agaaaggagacacgctgcggccctcgccgaagtgtgtcaaccggtgaagcctgtcaactcttc	3813
Dd	3952	TCCGAGATTAATGCAGAAATTATTTAGATCAAGTAGTACTGGGAAC--AATATCTTC	4008
OY	3814	gtgatgataaacatctgtctccaatatcgagaacgaatacgaagtgctcatctcgaacct	3873
Dd	4009	ACAATGATATAATCTCCTCGAGTTTGAAAATGAAATGATGTATATATTTAAACCT	4068
OY	3874	ctggcagcgtctctaagttcttccaagaacaacgctggagaaataacgtgtgaaccttaat	3933
Dd	4069	TTACCTGGAGATATATPAACCTTAAAAAAAACAATTGAAAAAACATTTTTCATTTAT	4128
OY	3934	gtcaacgtaaagacattctcgaaacgcgcgtttaatgaagagaataatcaagaagct	3993
Dd	4129	TTTTAAATTTGAACGATATCTTAAATTTACGCTTTTAAACAAAGAAAATATTTCTTACATGTA	4188
OY	3994	ttagagagcgaactgtatctccctaataagacctaccctcttaactacgtgtgcaagac	4053
Dd	4189	TTAGATCTGATTTAAATGCAATTTTAAACATATATCTCTCAATATGAATATCTTTGAAGAT	4248
OY	4054	ccatacaagttctccctaataagaagaagagataatcttgtctagttaacaatactt	4113
Dd	4249	TCATTTAAATATTTGAATTCAGAACAAAAAACACCTTTTAAAGAATGACAAATATATA	4308
OY	4114	aaggaictcatcgaccgccgttatcaatcttcgcataatgaatgtgtggtatgaagtc	4173
Dd	4309	AAGAGTACGTAGAAATGATATTTAAATTTGGCACAGGAAGTATAGTTATATGCAAG	4368
OY	4174	ctgacgaaaatacacaagcttgacacttgactcatlaaagaatlatacacgataag---	4230
Dd	4369	GTTTATAGGAAATATTAAGATGATTTAGATCAATTTAAAAAGTTATCCAAGAACAAAG	4428
OY	4231	-----caa	4233
Dd	4429	GAGAA GTTCCCATCATCACCACACACACCTCCGTCA CACGACAAAACACAGCACACAA	4488
OY	4234	ggcgagaatgaaaaatactgcgccttcctcgaataacatcgaaacctgtacaagcagtg	4293
Dd	4489	AAGAGCAAACTAAGCTTCCTTCCATTTTAAACAAACATTTGAGACCTTATACATATACTTA	4548
OY	4294	aacgacaaaatcgaccctcttcgtaattcacactgtgaagccaagttcctaactacttac	4353
Dd	4549	GTTATATAAATTGACGATTTACTTAATTAACCTTAAGCAAGATTAAAGCATTTGTAATGT	4608
OY	4354	gagaagagcaatgtggaagtltaaatcaagaagctgaactacctcaaaacaatccaagac	4413
Dd	4609	GAAAAAGATGAGACACATGTTAAATTAACTTAAAGTATGATTTAAAGCAATTTGATGAC	4668
OY	4414	aagctggaagtttccaagaanaaalanaaatctgcgtgaatgtgcagactgtctacgat	4473
Dd	4669	AAATAGATCTTTTAAAAACCATTAACGACTTGATGATCAATTTAAAAATTGTATANTAT	4728
OY	4474	tataaccaacaacatctcccgaccgaagtttctgtccaactgcatgtgtgtcgaaaactc	4533
Dd	4729	GATGCGAAAAAAGATATGCTTTGGCAATTAATCTTAGTACAGATAG--TTCAAATTTT	4785
OY	4534	gocaaacagttgctgagcaatctgtctogacgcaacctgcagggcatagtctgaactccc	4593
Dd	4786	CCTAATACAAATATATCAAAATTAATTTGAAGAAATTTCCAAGATATGTTAAACATTTCA	4845
OY	4594	cagcaccgaatgtgtgaaagaacagltgcgcccaagaatagcgcgtgttcaaggcatctygac	4653
Dd	4846	CAACACCAATGCCGTAAAAAACAATGCCGAAAAATTCGTGATGTTTCGACATTTAGAT	4905
OY	4654	gagcgcgaaggtgcaggtgtcttcctcgaactacaacaagaagagatagaagtcgtgag	4713
Dd	4906	GAAAGAGAGATATTAATGTATTTATTTAAATTAACAAACAAAGAGTGATTAATGTGTATAA	4965
OY	4714	aaccnaaacccttacctgcaatgaaacaatgygcgggtgtgacgcgaatgtcttaatgcac	4773
Dd	4966	AATCCAAATCTCTACTTTTAACGAAATAATAGTGGATGTGATGTGCAATGCCAAATGTAC	5025
OY	4774	gaggaaagcagcgcgtctcaaggaaagaanaatcacatgctgaatgtactaagccgcactcc	4833

Accession	Sequence	Position
Dd	GAAGACGTTTCAGGTACCAACGGAAAGAAATACATGTGATGATGTCATAACCTTATCT	5085
Qy	4834 taccacactctcgacgggalltlltgctccagctcctaattccctggcaccctccctg	4893
Dd	5086 TATCACACTTTCGATCGTATTTCTCGACGTTCTCTACACTCTCTAGGAAATATCATTTTA	5145
Qy	4884 ctgctccctacatgcacatccgtatcacgtcattcaatagatgaag	4939
Dd	5146 TTAAATACATGTTAATATTAATACAGTTTCATTTAAATAAATGTAGG	5191

Search completed: October 27, 2001, 13:14:31
Job time: 14107 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Comogen Ltd.

OM nucleic - nucleic search, using SW model

Run on: October 27, 2001, 09:20:04 ; Search time 136.83 Seconds
(without alignments)
6807.065 Million cell updates/sec

Title: US-09-269-874-1
Perfect score: 4920
Sequence: 1 atgaagatcatatcttctt.....tattacagttcattaa 4920

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PCRTUS.COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/Backfilest1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	ID	Description
1	2698	54.8	5181	US-08-257-073-10	Sequence 10, Appl
2	151.2	3.1	3095	5231168-1	Patent No. 5231168
3	136.4	2.8	4766	PCR-US93-07261-10	Sequence 10, Appl
4	135.4	2.8	2223	US-08-257-073-4	Sequence 4, Appl
5	132.6	2.7	6152	US-08-973-462-1	Sequence 1, Appl
6	132.2	2.7	19124	US-08-487-826B-13	Sequence 13, Appl
7	131.8	2.7	5361	US-08-973-462-2	Sequence 2, Appl
8	124	2.5	19124	US-08-487-826B-13	Sequence 13, Appl
9	118.6	2.4	9636	US-08-323-170B-1	Sequence 13, Appl
10	103.6	2.1	5852	US-07-867-106-2	Sequence 2, Appl
11	102	2.1	8920	US-08-446-855A-1	Sequence 1, Appl
12	102	2.1	8920	US-08-446-855A-1	Sequence 1, Appl
13	101.4	2.1	3763	US-07-792-865D-1	Sequence 1, Appl
14	101	2.0	658	US-08-998-416-595	Sequence 595, App
15	97.8	2.0	837	US-08-998-416-288	Sequence 288, App
16	96.2	2.0	6243	US-09-056-075-1	Sequence 1, Appl
17	95.8	1.9	3926	US-08-731-722-1	Sequence 1, Appl
18	95.8	1.9	3926	US-08-731-722-2	Sequence 2, Appl
19	95.2	1.9	8457	US-07-991-867B-1	Sequence 1, Appl
20	95.2	1.9	8457	US-07-991-867B-1	Sequence 2, Appl
21	94.4	1.9	1395	US-08-544-332-1	Sequence 1, Appl
22	94.4	1.9	1395	US-07-991-867B-25	Sequence 25, Appl
23	94.4	1.9	1395	US-08-107-755A-25	Sequence 25, Appl
24	94.4	1.9	6768	US-08-544-332-25	Sequence 25, Appl
25	92	1.9	636	US-08-107-755A-1	Sequence 1, Appl
26	90.8	1.8	615	US-08-998-416-1137	Sequence 1137, Ap
27	89.8	1.8	1511	US-07-991-867B-8	Sequence 8, Appl

28	89.8	1.8	1511	1	US-08-107-755A-8	Sequence 8, Appl
29	89.8	1.8	1511	2	US-08-544-332-8	Sequence 8, Appl
30	88.8	1.8	5852	1	US-07-867-106-2	Sequence 2, Appl
31	87	1.8	3933	2	US-08-731-722-3	Sequence 3, Appl
32	86.4	1.8	4673	5	US-07-638-431-1	Sequence 1, Appl
33	86.4	1.8	4673	5	PCR-US92-00018-1	Sequence 1, Appl
34	84.6	1.7	1511	1	US-07-991-867B-8	Sequence 8, Appl
35	84.6	1.7	1511	1	US-08-107-755A-8	Sequence 8, Appl
36	84.6	1.7	1511	2	US-08-544-332-8	Sequence 8, Appl
37	84.6	1.7	51952	3	US-08-947-823-1	Sequence 1, Appl
38	84	1.7	1230	4	US-09-140-466-1	Sequence 1, Appl
39	82	1.7	7218	1	US-08-232-463-14	Sequence 14, Appl
40	80.8	1.6	665	2	US-08-883-795A-36	Sequence 36, Appl
41	80.8	1.6	2447	2	US-09-014-969-14	Sequence 14, Appl
42	80.4	1.6	3701	4	US-08-845-258-10	Sequence 10, Appl
43	80.4	1.6	3701	4	US-08-990-571-10	Sequence 10, Appl
44	80.2	1.6	8700	2	US-08-392-625-16	Sequence 16, Appl
45	80.2	1.6	8700	2	US-08-466-961A-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-08-257-073-10
; Sequence 10, Application US/08257073
; Patent No. 576597
GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: de Talsne, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT FOXYVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; APPLICATION NUMBER: US 07/652,305
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,183
; FILING DATE: 20-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2570
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELETYPE: 425066 CURTMS
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5181 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-257-073-10

OY	2002	aaaaaagaalttgataaattaaagaatggttcaatgacctaaagttgaatcattgaataatgaa	2061
Db	2098	AAAAAAGAAAGTAGTAAATTAAAAAGATTATATACAAAAGTAAAAAGACATGTTAAAGAAA	2157
OY	2062	gaaaaaaaaacctaazaaacgaagttcaatctgataatcttggaacatcaaccgaaga	2121
Db	2158	GAACAAGCTGCTTATACAGATTATACACAACCTTTAGTTGCGACGAAGCGAAGC---AACT	2214
OY	2122	gaataacaagagacaagcaactacaaaaccctgagcaacaagcaggtatcttccttagaaga	2181
Db	2215	GAAGATGGGGGTCACTCCACACACACACTTATGCCATTGAGGAGAAACAGAACTAACACAA	2274
OY	2182	gattcagtagacaagacaagacaagaacaanaaacaagcaacaaccacagttacagtaacca	2241
Db	2275	GAAACAGAGAAAGAAACGAAAGAAACAGTAGAGCACACAAACAGCGTAAACAATTAACATTACCA	2334
OY	2242	gtaccagaagaagaagacacaagttcccaaccacccagcccgtaataataaactgaa	2301
Db	2335	CCAAAAGAAAGTAAAGTTGTTGAAAAATTCATAGAACATAGAGATTAATGCATTAATCCACA	2394
OY	2302	aatgtttcccaacttagatactcttgaanaaattatagaaattttaaactcatatata	2361
Db	2395	GCCCTTGACAAAAACAGTTTACTTAAGAAATTTAGATGAATTTTAACTAATCATATATTA	2454
OY	2362	tgtaacaatatatttggtttcacactcaactatgaacgaagaagataltaaaacaatat	2421
Db	2455	TGTCATTAATATATTTATGATCAAACTGATAGATGACCAAAATTTATAGAGTATAT	2514
OY	2422	aaattacaaagaggaagaagcaaatatagttcatgtgacatcattagactatgttt	2481
Db	2515	AATCTTACCTCCAGAGAAAGAAATGAAATTAATCAATCATGATGACCAATTTGATTTATATTT	2574
OY	2482	aattacaaataacactacacttgtaagtatctatgttttgatagotaaacaatagtta	2541
Db	2575	AATATTCAAAATTAACATACCTGCTATGATTTCTTATATGATAGTAGTACAAATCAATTA	2634
OY	2542	tcacaactattatlggaatttgaanaaagaatlggtttgaatttatalaacttaag	2601
Db	2635	CAACATCTCTTTTTGAAATTAATTCAAAAGAAATGATTTATTTATTTACATACTAAA	2694
OY	2602	gataatgacaaatltlaaaatttattatagaggaaacgaaaaa-----gltacacactc	2655
Db	2695	GAGGAATATCAATCAAAAATTTATTAAGAGGACAAAACAAATTAACGGAACATCATCT	2754
OY	2656	gtaaaaactcttcaagtatcaatgaagcaacatatcaatgaaccctcagagataaacc	2715
Db	2755	ACATTCACAGTCTGGAATTAACACCGTAATTAATGCTCAATCCGCAATCCGACATCGATTAATCC	2814
OY	2716	ga-----	2717
Db	2815	CAAAACCAACATCAATTCAGCTCTGACCATTACCCAAAATGATGTAGCTATCATCT	2874
OY	2718	-----agtaatgcaaatgatatgatacaatcactcaaatltgaataatagt	2766
Db	2875	GGTCCCTGCTGTAGTTGAAGAAAGTCAATGATCCCTTAACAGTATTTGCTATTATGTAACGAT	2934
OY	2767	ltaaattatttgaanaacatattgagctcttggaanaaacaanaatatataaccagaatta	2826
Db	2935	TTGAAAGGTATGTGTAGTCTCTTAAATCTTGGAATTAACACTAAAGTACCTAAATCCATTA	2994
OY	2827	---atagtcacaaaaagtagtgaanaacttltgaagaagataltaaaagatagtgatata	2883
Db	2995	ACCATTTCTACCAACAGAGATGAGAAAATTTTATGAGAAATATTTTAAAAAATATAGATAC	3054
OY	2884	tttataatgaatccttttacaatttttgaanaacttaaaagctgatgatataatcattg	2943
Db	3055	TATTTTATATGATATATCAACAACTATTCGTAATAATCTAAATCAAAAGTAAATTAACGGTTTG	3114
OY	2944	aatgaatgaacaaagaagaaatattgagaagaagatataataaattbaaaaaactta	3003
Db	3115	A---CCGAAACACAAAAAATGCATTTAAATGATGAATTTAAAAAATTTAAAAAGATACATTTA	3171
OY	3004	cagttacatttgattatataataataataatlaaatlaaatagaagaatttcttgataaa	3063

Db	3172	CGTTATCATTTGGATTATTATTAATAAATTAATAATTAATAATTAAGATGATTTATTAATPAAG	3231
Qy	3064	aagaaacagcttggtaaatataaaatgcaaaataaaacttacttataaagaacaa	3122
Db	3232	AAAAAAGAACTGGCCAGACAAAATGCCAAATTAATAAAACCTTATTATAAAGACAA	3291
Qy	3124	ttagatccaattgaatttcacttaatacccaagcagtatacaaaacttcctc	3183
Db	3292	TTAGAGATCAAAATTCATTCCTTAATTAACCCCATATATGTATTACAAAACCTTTCTGT	3351
Qy	3184	ttccttlaacaaaataaaagaaagcttgaatlagcgaacttgaataacacatlagaataaca	3243
Db	3352	TTCTTTAACAAAAAAGAAAGCTGGAATFAGCAGAAATGGAACATTAAGAAAACAA	3411
Qy	3244	aaaataatttgaacacatlalaaagagcttgttaaatatataatggtgaatcatctca	3303
Db	3412	AAAAATTTATGAAACATTAATAAAGGACTGTTAAATATTATTAATGTCMAATCATCTCCA	3471
Qy	3304	ttaaaactttaaagttgaagatccaattcaacgaagatataatgcccgtttgaagaac	3365
Db	3472	TTAAAACTTTAAGTGAGATATCATTCATTCAAACAGAAATATATATGCCAAATTACAAAAA	3531
Qy	3364	tttaaagattaaagtlaaaatagaagaaaaatlaaagataattlaaattlagaagaaga	3422
Db	3532	TTTAGAGATTAAGTAAGTAATAATGATGGAACCTCAATGATATATTACATTAGAAAAGAA	3591
Qy	3424	aaattatcaacttatacgaatgagatcatcatcttaattgtctgaattaaagaatlaa	3483
Db	3592	AAATTTCTTTCTTATCAAGGGATTCATCATTTATTACTGAATTAATAAACAATTAATA	3651
Qy	3484	aaaaataaaatttacaaggaattcccaagtgaaataataacgaatgtatacaatgca	3543
Db	3652	AAAAATTAATAATTTACAGGTAATTCCTCAAGTGAAATAATPATAGAAATTAACGAACCT	3711
Qy	3544	ttagaatcttcaaaaaaatttccccaagagagacagat-----	3582
Db	3712	TTAAATTCCTTCGGAATTTTCTCCCGAAGCAAAAGTTACAAACAGTTCTAATCTCCACT	3771
Qy	3583	-----gtlgaacagctgttaagtgaagatgagatccgaacatlagaa	3624
Db	3772	CAACAGATGTAACTCCATCCATCCCATTTATCTGTAAAGGTAAAGTGTGACAGATCCACA	3831
Qy	3625	caagtcacccaagaagaacacagatcaactcatgtatggagccagatctacaacatlaa	3684
Db	3832	AAAGAGAAACACAATTAACCAACTCCAGCTCTTTTAAACGAATTAACCAACTACTTA	3891
Qy	3685	acatcacaaaatgccccgttgaatgaatgaatgaatcatatagtaactatcttggagaa	3744
Db	3892	CAATTACAAATTTATGACGAAGAAAGATGATCTCTTGTTGTTATTCACCTTTTGGAGAA	3951
Qy	3745	tcogaagaagattatgatattlagaacaaatgataacagagaagcagtaactcctcc	3804
Db	3952	TCCGAGATATATGACGAATTTTATGATCAAGTATGTAACGTGGAAGACATA---TCTGTC	4008
Qy	3805	gtaattgtaacacttcttcaaatlgaanaatgaatagtaggttttatattltaaaact	3864
Db	4009	ACAATGGATATATATCCCTCAGATTTGAAAAATGAATATGATCTTATTAATTTAAACCT	4068
Qy	3865	ttagaagctgttllatagaagtllaaaaaaacatlgaanaatlaaoglatagacttcat	3924
Db	4069	TTAGCTGGAGATATATACACTTTAAAAAACAATTTGAAAAAACAATTTTATACATTTATAT	4128
Qy	3925	gttaatgtlaagatattttaaattcaacgaattatlaaaacgtlgaanaattlcaaaaatgt	3984
Db	4129	TTAAATTTGAACGATATCTTAATTAATTCACGCTTAGAAGAAAGAAATATTCTTGAGTGA	4188
Qy	3985	ttagaatcaagtttaattcatataaagttttaaactcaagaatlatagttgtcaagat	4044
Db	4189	TTAGATCTGATTTAAATGCAATTTAAACATATATCTCTCAAAATGAATACATTTATGAGAT	4248
Qy	4045	ccatataaattcttcaataaagaaaaagatlaaatttccaaagcagtataatataatt	4104


```

Db 4249 TCATTTAAATTTGTAATTCGACCAAAAAACACACTTTTAAAAAGTTACAAATATATA 4308
QY 4105 aagattcaatagatcgcgatafaaattttgcaaatgctcttgatattataaata 4164
Db 4309 AAGGATCAGCTAAGAAATGATTTAAATTTGCACAGAGAGTATAGTTTATGAAAAG 4368
QY 4165 tlatcggaaaataataatagaattagattcaattaaaaataataca----- 4214
Db 4369 GTTTTGGCCGAATATTAAGATGATTTAGATTCATTTAAAAAGTTATCAAGAGAAAAG 4428
QY 4215 -----cgacaacaa 4224
Db 4429 GAGAGTTCCCATCATCCACCAACACACCTCCGTCACACGAAAAACAGAGCAAA 4488
QY 4225 gctgaagaatgagaataactccctcttttaacaatatagacactataaacaagtc 4284
Db 4489 AAGAGGAAGTAAAGTTCTTCCTTTTAAACAACATTTAGACCTTTATACAAATACCTTA 4518
QY 4285 aatgataaaattgattatattgaatcattagaagcaaaagttcaaatatatacat 4344
Db 4549 GTTAATTAATTTGACGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4608
QY 4345 gagaatcaaatgagaagtataaataaagaacttaattacttaaaacaatcaaac 4404
Db 4609 GAAAGAGATGAACACATCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4668
QY 4405 aatttgcagatttaaaaaaataacaatttgcgtggaattgctgattatcaaacag 4464
Db 4669 AATATAGATCTTTTAAAAACCATTAAGCACTTGAAGCAATTTAAAAATTTGATTAATGAT 4728
QY 4465 tataacataataactattgacaaagttccttagtagaacgtagtatttgaataatc 4524
Db 4729 GATACGAAAAAATATGCTTTGGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4785
QY 4525 gctaaaaacgtttatctaatcttgaatgagaacttgcgaagtgatgataaacattca 4584
Db 4786 CTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4845
QY 4585 caacacaaatgcgttaaaaaaacaatgctcaaaaattcgtgattcagacattagat 4644
Db 4846 CAACACCAATTCGTTAAAAACATGTCAGAAAATTTCTGAGTGTTCACACATTTTGAT 4905
QY 4645 gaaagagaatgtaaatgatttatttaataatacaaaagaagtgataaagtgtgaa 4704
Db 4906 GAAAGAGAAATGTAATTTTATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4965
QY 4705 aatccaaatcctactgttaacgaaaataatgctggtgtagtgcagatgcacaatgtacc 4764
Db 4966 AATCCAAATCTCTCTGTTAAAGAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5025
QY 4765 gaagaagattcaggttagcaacgcgaagaataatgacatgtagtatttaaacctgattc 4824
Db 5026 GAAAGAGATTCAGGTACCAACGAGAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5085
QY 4825 tatccacttctgattgatttctgcagttccttaacttcttaggaatatacttca 4884
Db 5086 TATCCACTTTTCATGATGTTTCTGCAAGTTCCCTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5145
QY 4885 ttaataactatgttaataataatacaagttcaattaa 4920
Db 5146 TTAATCTCATGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5181

```

RESULT 2

```

5231168-1
; Patent No. 5231168
; APPLICANT: DZIEGIEL, MORTEN, BORRE, MARTIN, JEPSEN, SOREN,
; VUUST, JENS, RIENECK, KLAUS, WIND, ANNETTE, JAKOBSEN, PALLE H.
; TITLE OF INVENTION: MALARIA ANTIGEN
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/409,658
; FILING DATE: 18-SEP-1989

```

```

; SEQ ID NO:1:
; LENGTH: 3095
5231168-1
Query Match 3.1%; Score 151.2; DB 6; Length 3095;
Best Local Similarity 42.9%; Pred. No. 4,5e-16;
Matches 1310; Conservative 0; Mismatches 1688; Indels 56; Gaps 9;
QY 512 atgaagaatlaataatgaattatataataaacttttttggatttttaagcga 571
Db 47 aagaagtagtcttgaagaagggccatcaagaagaatttgaacttgaacaataaagaag 106
QY 572 aattaaatgtagttagtgcgaatgatttgcgaatcctcaatcttaaatctg 631
Db 107 aatcaggtgaagaatgaattatgtgataatgaagaagtggttttgaagaagccatcag 166
QY 632 caaatgaattagacgtacttaaaaacttggttcggtatagaagaacattagaacata 691
Db 167 aaattttcatctlgaagtagtaactctgaatttaaatgaagaattggttgaatc 226
QY 692 ttaagaataatgagaagaatggaagattacattaaaaaataaaaaacataga 751
Db 227 acaaaagtgaactgaacccgtcgtgaacatgaagaagttgattcgtgaagaagaaccc 286
QY 752 atataatgaattaatgaagaagaatgaagaacaattgataaataaagaatgcaacta 811
Db 287 aaccagctgaagaatgaagaagaatgtagtaagaagctctcgtgaagaatgttact 346
QY 812 aagaagaagaagaagaagaatlaacaaagctcaataatgattcttcaataaataac 871
Db 347 aacaaatgtagaagaatcagtggaagtgattgattgattgaagaagtgattt 406
QY 872 aattagaagaagcacataatttaataagcgttttagaagaacgtattgaactttaaa 931
Db 407 aagaacctaalcagaaatltgaacgttgcataaatgactcgtgaattgaatgaatga 466
QY 932 aaaaatgaacatlaaggaatctactgataagaataatgaataaaccctccccc 991
Db 467 aattagttgattcagaagaagaatgtagtaccagctgacatgtgataatgttact 526
QY 992 ccaattctggaataacacccaatactccttgatagaagaacaaatcgaagaacg 1051
Db 527 aaaaagtgatctgacagcagctgaacacgtagaatgttactcgtgaagaagtagac 586
QY 1052 aaaaagaataaagaagaatgccaacataatlaattatattgaattgaattact 1111
Db 587 aaccagctgacatgtagaagtgatctgaacaaatgaataacgaacacccgaaga 646
QY 1112 atccactgattagaataactatttaagaagaagaataaataatgataatgaag 1171
Db 647 aagaatgaccaggtctcctcaaaacattgaagaagaatgaagaagtgatgacacta 706
QY 1172 aggttgaacaaagaatcaactgaaccccaatgaatatccaatggaatgattactc 1231
Db 707 aattgtagaccttcaataatattgaacctaatlttggctcacaacacaaatccaa 766
QY 1232 tgcataatacgaatataaataatgctttaaataatgaacttaattcttgggtatt 1291
Db 767 aaccagttgacacatcatltgtcaaaaatltgaagaatltcctctgagaagaataa 826
QY 1292 atcatttgaattatacaagaagaacgaatgaagaacattatctgataatgaagaaga 1351
Db 827 caagtgtgacctggaagtagaagaagaagaatgatactggaagtgttgaagaagaac 886
QY 1352 aattcataatgaatlaagaagaagaatlaaataagaagaagaagaagaatgactgata 1411
Db 887 aaatcacagaatcagttgaagaagaatccagtaaatggaagtgatttgaagaagttc 946
QY 1412 aaaaatcttaagaagacagatcgaagttttaaataatgaatacaagaagaatgata 1471
Db 947 acactgaacattagattttagatcataaagaacagtgatccaagaatagtagaagttga 1006

```

1472 tacttaatgaaatttatagtacaaattccaataataatagatttaacttaatttcgaaa 1531
1007 aaattcctcaagaactacatgaaatgaagtgtgtcattccagaatattgttgaattgag 1066
1532 aaatgatggtaaaagatatctcatataag---tgaagaacttacacacataactt 1588
1067 aagttttctcgaaccaataacaaatcaacgaatttccaagaatttaagtataaaa 1126
1589 tgcataccatgaanaattccaataactctgaaagtttaacaaaagctcttaata 1648
1127 gtcgcatacttcgaatgaatagtagaagtagaagaatacttccaagaatgataaaa 1186
1649 tgaagaattattcttaagaatatagtagtgaagaagaattaaataataataaatt 1708
1187 atgaaaaagttgaacatgaaatagtagaagtttgaagaattctaccagaagataaaa 1246
1709 taataagcaaaatagaaa-----atgagattgaacatagtgtgaataatata 1756
1247 aaaaagtcacatgaataagtagaagttgaagaattctaccagaagatgataaaa 1306
1757 aaaaagatgaagaacagct-----tttgaaaaaaaattacttaagcgcgaataaac 1810
1307 aaaaagttgaacatgaataagtagaagttgaagaattctaccagaagataaaa 1366
1811 cgaatgaaaaaatttgaagaatctgcacattgtaaaagtacaagttcaaaaagtttat 1870
1367 aaggtcaacatgaataagtagaagttgaagaattctaccagaagataaaaatgaaa 1426
1871 taatgaacaaattgaagaatlaaaaagatcaattgattttaaanaaattgataaa 1930
1427 tgcacatgaataagtagaagttgaagaattctaccagaagataaaaatgaaaagctc 1486
1931 aacataatactatcttcccaattcttacaacagaanaataagcagaacttatatt 1990
1487 aacatgaataagtagaagttgaagaattctaccagaagataaaaatgaaaagttcaac 1546
1991 taattgtgtgtaaaaaaagaatttgataaattaaagtgtcttcgaagtagaatac 2050
1547 atgaataagtagaagttgaagaattcttccagaagataaaaatgaaaagttcaacag 1606
2051 tgaataatgaagaaaaaaacataaaaaacagaaggtcaatcogtaattcggaaacat 2110
1607 aaatagtagaagttgaagaattcttccagaagataaaaatgaaaaggtccaacatg 1666
2111 caaccgaagaagaataaagaagaacgaactcaaaaacgttgacagaacagcagctcg 2170
1667 aatatagtagaagttgaagaattcttccagaagataaaaatgaaaagttccaacatgaaa 1726
2171 cttagaaggaagattcagtagacaagacaagacaagaacaaaacagcacaacccag 2230
1727 tagtagaggttgaagaattcttccagaagataaaaatgaaaagttccaacatgaaatag 1786
2231 taccagtagcagtagacagaagcacaagtcaccaacaccacagcagaacgaataata 2290
1787 tagaagttgaagaattcttccagaatgtgtgaattgaagaagtagcaccatcaacaaca 1846
2291 ataaagctaaaaatttcccaattagattatcttgaanaattatataatgattttaata 2350
1847 ataaacaatgaataatgaacataaaaacagaagaacaaaagaattgaatttagtgtcg 1906
2351 ctctacataatgtaacaaatatttggtttcacacccaatgaacgaagaataat 2410
1907 aagaaaagaacattccacaagaaccggtgtacctcatttaaatgaaaatgaaaagctta 1966
2411 taaacaataataaattacaaaggaagaagaacgaataatgaattgtagcattag 2470
1967 ctccaacaacatctgaaggtgattccactaaaccagatagttccaattaaatagtag 2026
2471 acttatgtttaataatacaaaaataacatccgttaatgattctatgattgtagcttaa 2530
2027 aagaaaataaacaataaagaagaacaccagtagtagtgcacaaaacatgtagaac 2086
2531 acaatagttatcacacactatttatgaaattatgaaaagaatggtttgtaatttat 2590

2087 aaaaatacagaagaatgataatgataagaagatgatagtatagtttgaaggt 2146
2591 ataacttaaggataatgcacaaaataaaaatttatttagagaagcgaanaaagtcca 2650
2147 tatcaagaagaatgataagaag-----atcattcaataaataaataaagaatactat 2200
2651 catctgttaaaccttcttaagttcatcatgcaaccttatcatatcaacc----- 2704
2201 ctcttaacataatatactacaagaagaatttlaaaaagatctccaactatgtgaatg 2260
2705 -----aggataaccgaagtagtgaagtcgaatgatcatcatcatcacaatttga 2758
2261 ttatgttaatgataatgataatgatttccaaagtgtgaagtacaaatlaagaataagca 2320
2759 ataatagtttaaataattatttgaanaacatgtgagttggaanaaaacaaatataacc 2818
2321 aggataatgaataatttccaaaacatttaaaatlaatacaaaaaaaataaaaaaa 2380
2819 aagaattataggtcaaaaagtagtgaaaacttt-----atgaaaagatatlaa 2869
2381 aaaaaaaaataataaataatttcttcttataatagtaactgaattatattat 2440
2870 aagatagtagatacttlaaatgaactcttacaacatttgaacattcgaagctgag 2929
2441 aataaataatataatataataaagaatactagagtttctgtatataatagaataaattc 2500
2930 ataataatcaatgaatgaatcaaaaagaagaatlaagaagaatlatbaataat 2989
2501 atgtatattatatagaanaaaataaataaataattcttcttctagtgtaataat 2560
2990 taaaaaaacttaccagttcatcatgttatatatataataaataaataaattgaaa 3049
2561 actctaaataataaataaataaataatagaacatttaccagatatttataatctcgaaa 2620
3050 gattattgtataaagaagaacaggtgtgtaaatataaagtgaacaaatlaaaaacttact 3109
2621 gatacatgtataaanaaaatttctgcg----ttataataataaataagaanaattatca 2677
3110 tattaaagaacaaatagataccaanaattgaattcacttaataacccaagcatgtatcc 3169
2678 tttaatgataatttlaagaanaatatataatataatgataatgaagaagaanaaaa 2737
3170 aaaaacttctgttcttcttaacaaaaaaagaagctgaatatagaagaactgaaaaca 3229
2738 aaaaatttcttatacatatttcttcaataacatataaataatataatataataata 2796
3230 catlagaanaacaaaataatattatltgaacatataaaggaactgtttaaataatag 3289
2797 ctggaataaattattatgataataattataataacacaggtgtgcttggcataaatgc 2856
3290 gtaatcatctccatlaaaaacttgaatgaagaatcaattccaacagaagaataattag 3349
2857 aataaaaataltgaacaataaataatcaltgt-----tctgattatattcttaggttaaaa 2912
3350 ccaagtttagaaaacttgaagtagtaatttagaagaagaataatgaagaatattaa 3409
2913 ttcataataatttttttaattttaaattttaaacaagaatgaataatlaaataaa 2972
3410 atttgaagaagaanaattatacatatacttcaagttagatatacatcattaatgtcgaat 3469
2973 taagaaaagaatgaatttttaataatgttcataaatttttttttttagtagaataat 3032
3470 taaagaagtaataaanaataaataatatacaggttaattctccaagtgaanaata 3523
3033 ttaaaataataatataatlaagttatattatagtaactattttagtaataata 3086

RESULT 3
PCT-US93-07261-10
; Sequence 10, Application PC/TUS9307261
; GENERAL INFORMATION:
; TITLE OF INVENTION: PFEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF

Db	1660	GAGATNTCAACTTGGATATGAAGGTCCTCTACATTTAAAGAAATATGATCAACAGAAAT	1719
Oy	2012	ttgaataatlaaagaatglttcacgtcctaagtgaaatcaltgtataaetgaagaaaaaa	2071
Db	1720	TAGCAAAAGGTAAGGATATATACAAACACACAGAAATCCGTGGATATATGATCAAA	1779
Oy	2072	acataaaaaacgaagtcacatcggaataaltcggaaacatcaacggaaggaataacag	2131
Db	1780	CAGAAATTAGCAAAAGGTAAGGATATATACATACAAACACCATGAAATCCGTGTGAAATATG	1839
Oy	2132	gacacgaacatlaaaaaacccctggacacaacgagatctgctttagaagagatcagtaac	2191
Db	1840	ATCAATATCAGAAATTAGCAAAAGGTAAGGATATATACATACAAACACCATGAAATCCGTGTG	1899
Oy	2192	aagcacagacacagaacacaaaacaagcacacacacagtaaccagtaaccagtaaccagag	2251
Db	1900	AATATGATCAACAGGAATTTAGCAAAAGGTAAGGAACTACTATATTAACACACGAAATC	1959
Oy	2252	caaaagcacaaatgccaaacccacccagcacagtaataataaacctgaaaatgtttcca	2311
Db	1960	TGGAAAGATACAAATGAACATGATTTAGCAAAAGGTAAGGAGTACTATACAAACCCATCG	2019
Oy	2312	aattagaattccttgaaaaaattatgaattttaaactctcatatataatgaagaaat	2371
Db	2020	AATCCGTTGATGAATATGATCAATACGAATTTAGCAAAAGGTAAGGATATATACATCAAC	2079
Oy	2372	ataatttggttcaacctcacaatgaacgaagaagataltaaacatatataaaattacaa	2431
Db	2080	CACACGAATCCGTGGATGATATGATATCAACAAAGAAATACCAACAGTAAGAAATACATA	2139
Oy	2432	agagagagaagaagcaaatgaatgltcagtgtatccattagaactatgtttaataacaa	2491
Db	2140	ATAAGACACGCAAAATCTGAAAGAAATATATGAACATGATTTAGCAAAAGGTAAGAG	2199
Oy	2492	ataacataactgtaatgatatctatgtttgataagcttaaacatagttatcacacat	2551
Db	2200	TAACTATTAAGCACCGCAAAATCTGAAAGAAATATATGAACATGATTTAGCAAAAGGTA	2259
Oy	2552	ttatgaaattatlgaaaagaagaatgltctgtaattatataaactgaagataagca	2611
Db	2260	AGGAATGATCAATATTAAGGCCATGAAATCTGGAAGAAATATATGAACCGATTTAGCA	2319
Oy	2612	aaatlaaaaaattatagaagaaagc-----aaaaagtalccacatctgtlaaaactctt	2668
Db	2320	AAGGTAAAGGAATGAACTTAATTAAGGCCATGAAATCTGAAGAAATATATGAACATCTGAT	2379
Oy	2669	caagttcatcaatgaacccatcatcatcaaacctagaataacccgaagaaagtga	2728
Db	2380	TAGCAAAAGGTAAGGAATGATCAATTAAGACACATGAAATCTGAAGAAATATATGA	2439
Oy	2729	atgatacatcatcactctcacaaattgtaatatagtttaaatatatttgaaacatat	2788
Db	2440	CTGATTTAGCAAAAGGTAAGGAATGAACTTAATTAAGACACCGCAAAATCTAGAAATAT	2499
Oy	2789	tgagltctggaaaaaacaaatatataaccagaatlaaiaagtlcaaaaaagttagaa	2848
Db	2500	ATGAACATGATTTAGCAAAAGGTAAGGAATGAACTTAATTAAGACACCGCAAAATCTAG	2559
Oy	2849	acttlatgaaaagatatataaagatagatgatacattataaagaatcctttacaat	2908
Db	2560	AATTAATTAAGAACTGATTTAGCAAAAGGTAAGGAATGAACTTAATTAAGACACCGCAAA	2619
Oy	2909	ctgtgaatcctaaagctgagatcat-----taattcatgtgaatgataatcaaaa	2959
Db	2620	TAGAAAGATATATGAACATGATTTAGCAAAAGGTAAGGAATGAACTTAATTAAGCACCG	2679
Oy	2960	ggaagaatgaagaagatatataaataataaaaaaacttacaatlatcatattgtt	3019
Db	2680	AAATATGAGAAGAAATATATGAACATGATTTAGCAAAAGGTAAGGAATGAACTTAATTA	2739
Oy	3020	tataataaataataataaattcgaagaatattctgataaanaagaacagctgtgta	3079
Db	2740	CACCGCGA-AATATGAGAAGAAATATATGAACATGATTTAGCAAAAGGTAAGGAATTAAT	2798

OY	3080	aataataaaygcaaatlaaaaaaacttacttattlaaagaacaattagaaaccaaatytga	3139
Db	2799	AATTAAGACCCCGAAATCTAGAAAGATATGAAAGAAAGAAATATATGAAATAT--G	2856
OY	3140	attcaacttaaacaccgaagcagctatlaacaaacttcctcgttcttcaacaanaa	3199
Db	2857	AATTTCACAAATTAAGAGTCTGATGATTTAAAGAAATGCAAGCTAAATAATTAAGAT	2916
OY	3200	aagaagctgaatatgcaagaactgaaacacatlgaaaaacaanaalatattgaaac	3259
Db	2917	TACGAAATTAAGCATCTGATGATTTAAAGAAATGCAAGCTAAATAATTAAGATATAC	2976
OY	3260	atbtaagaagctgtttaaattatataatgagyaatcattccattaaanaacttaagt	3319
Db	2977	GAATTAAGGCTCTGATGATTTAAAGAAATGCAAGCTAAATAATTAAGATATACGA	3036
OY	3320	aagaatcaatcaaacagaagaatattatgcccgttltgaaanaacttlaagatlaagta	3379
Db	3037	ATTAAGGATCTGAAGGATTTAAAGAAATGCAAGCTAAATAATTAAGATATACGAATTA	3096
OY	3380	aattagaagagaataatlaagyaataattlaaattlagaanaagaaaatattcatcattat	3439
Db	3097	AAGGATCTGAAGGATTTAAAGAAATGCAAGCTAAATAATTAAGATATACGAATTAAG	3156
OY	3440	caagtggattcatcattatcttcgttgaatlaaagaagtaataanaataanaattla	3499
Db	3157	GATCTGAAGGATTTAAAGAAATGCAAGCTAAATAATTAAGATATACGAATTAAG	3210
OY	3500	caggtaattcccaagtgaaataataacagatgttaacacatgcatlagaattcaataa	3559
Db	3211	AAGGATCTGAAGGATTTAAAGAAATGCAAGCTAAATAATTAAGATATACGAATTAAG	3270
OY	3560	aatttccccaagaagagacagatgttgcacaagttgtaagtgaaagtgatccgcacat	3619
Db	3271	GATCT-----GAAGGATTTAAAGAAATGCAAGCTAAATAATTAAGATATACGAATTA	3324
OY	3620	tagaacaagatcaaccgaagaacacagcatlcaatcatgtlaagagcagagtlcaacaac	3679
Db	3325	AAGGATCTGAAGGATTTAAAGAAATGCAAGCTAAATAATTAAGATATACGAATTAAG	3384
OY	3680	taacaacatccaaaagtctgatgtagaagtagagtagagtaactagtaactatttg	3739
Db	3385	GATCTGATGATTTAAAGAAATGCAAGCTAAATAATTAAGATATACGAATTAAGAT	3444
OY	3740	gagaatccgaagaagatcatgatgatttagagcaaaatgtaacagaggaagcagtaactc	3799
Db	3445	CTGATGATTTAAAGAAATGCAAGCTAAATAATTAAGATATACGAATTAAGATATG	3504
OY	3800	cttcgcgaattgataacatactcttctaaaaatgaaaaatgaatataggtttataattaa	3859
Db	3505	AAGGATTTAAAGAAATGATATATACAAATATGATTTAAAGATTAACGATTTCAATA	3564
OY	3860	aacctttagccaggtgttataagaatttlaaanaaacatltgaabaataagctatgacat	3919
Db	3565	AAGATTTATCTAATAAAGACATGAAAAATTAAGACATATTAACAACAAGATATTCTTA	3624
OY	3920	ttaatgttaatgtlaagagatatlttaaatccagattataaacytgaanaattcaaaa	3979
Db	3625	AAGCATATGAACAAATAAAGACATTTAATAATTAAGATTTGCAATGAAACATGAAGATTA	3684
OY	3980	atgttttagaatcgaatttatattccatatagaagatttaacatcaagtaattatgttgca	4039
Db	3685	AAGACATTTAATAATTAAGACATTAAGATTAAGACATTTGAAAGTATAGAAACATGAGAC	3744
OY	4040	aagaatccataaatttcttcaataagaanaaagaagataaattctttagcagttlaatt	4099
Db	3745	AACAAATATACGCTTTAAAGATATACACTAGTTAAAGACACAAATATACAGSTTTAAAG	3804
OY	4100	ataatgaagatccaatagatacagataataatttgcacaaatgagatgtcttgatataata	4159
Db	3805	ATACACCTTAATGAAGAACCAACAAATATACGTTTAAGATTTACACCTAGTGAAGACAC	3864

```

OY 4160 aaatattatccgaaataataatcagatttgatcatcattaaatatatacagaca 4219
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3865 AAAAATACGGCTTTAAAGAAATACACCTAGTGAAGGACACAAATACTGTTTAAAGATA 3924
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 4220 aacaagtgtaaatgagaata 4241
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3925 CACCTAATGAAAGACACAAAA 3946
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
US-08-257-073-4
Sequence 4, Application:us/08257073
Patent No. 5766597
GENERAL INFORMATION:
APPLICANT: Paoletti, Enzo
APPLICANT: de Taisne, Charles
APPLICANT: Tine, John A.
TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue, 25th Floor
CITY: New York
STATE: New York
COUNTRY: UNITED STATES OF AMERICA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/257,073
FILING DATE: 09-JUN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/075,783
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/852,305
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/672,183
FILING DATE: 20-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2570
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066 CURTMS
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2223 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-257-073-4

Query Match 2.8%; Score 135.4; DB 1; Length 2223;
Best Local Similarity 44.6%; Pred. No. 1.3e-13;
Matches 869; Conservative 0; Mismatches 1041; Indels 39; Gaps 7

OY 329 gtacaatccctcaataatcaagtatcagatgctaataatcctaagctgatttbaaac 388
      || || || || || || || || || || || || || || || || || || || || ||
Db 284 GTTTTAATTCACCAATATATTTCAGCAAAATTAAGTAAATAAATAACGTGTAAGAAATGA 343
      || || || || || || || || || || || || || || || || || || || || ||

OY 389 acagagtaacgaattactgttactatcatcaagaactca--aatacctcaactcttga 446
      || || || || || || || || || || || || || || || || || || || || ||
Db 344 AAAATTTAGTAATTTCTCATCAATCTCATTAAGAAATTAATTAAAGCATTAAGAAAGATA 403
      || || || || || || || || || || || || || || || || || || || || ||

```

```

Db 1472 TGAACAGTATTATAATGATGATGAATAATGATTCTTTGAAAAAGAAATTAATCGAAA 1531
Oy 1577 accataactcttgcacccctatgaaatctcaacataactctgaaagttacaaag 1636
Db 1532 ATAAATTAACACCTTAATGTTGACCAACCAATCAAAAAAAGAAATGAAGAA 1591
Oy 1637 cctttaaataatgaaagatacttcttaaggaataagtagttgaaagaatlaaat 1696
Db 1592 CTGTTCTCGTATGATGAATAATTTTGAATATCCTCGAAATTAATTTTAAAGAAAT 1651
Oy 1697 attataaataatlaaagaacaaatagaaatgaaatltgaaacattagttgaaatata 1756
Db 1652 ATTATTATGATGAATAATGATGATGAAGTAAAGTTAAAAAATAGGTGCACATTAA 1711
Oy 1757 aaaaagatgaa-----gaaagcgtcttltgaaaaaaaatctactaaagcga 1804
Db 1712 AAAAATTTGAACTTAAATAATGAAATGTTAGTGAACCTTAATTAATTTGATTCTT 1771
Oy 1805 ataaacagatgaaataatlttagaagtaactgacatgttaaagtaacaagttcaaaa 1864
Db 1772 GAAATTAAGATTAATAAACCATTTGAAGCTATTAACACAGATTTTCAATTTAATTAAC 1831
Oy 1865 tttaataatgaaacaaatltgacgaatlaaaaaagactcaattgttlaaaaaatgtag 1924
Db 1832 AATTACAAGCTATTATATGAATTTATGCAATTTATGCAATGGAACAAATAATTTCAAC 1891
Oy 1925 a---atlaaaacataatatacagtctcccaatcttacaacaaagaataaagaagac 1981
Db 1892 AATATTTCACAAATAATTTCTAGAAATGATGTTCTTATCATCAAGAACGGAGAA 1951
Oy 1982 ctattattattactgttctgaaaaaagaatltgataaataaagtgcttcagccaaag 2041
Db 1932 TCGAATAACAACTTAAGCAATTCACCAACCAATAGAACCTGAAGTGGCTTCGCAC 2011
Oy 2042 tagaatcattgataaagaagaaaaaaacataaacaagaagtcactcgataat 2101
Db 2012 CAAAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2071
Oy 2102 cggaaacatcaacggaaggaagaataacagagacgaactcaaaactgtgacaaga 2161
Db 2072 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2131
Oy 2162 caggaatctgacttagaagagagatcagtaacagcagaagacagaacaaagaagc 2221
Db 2132 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2191
Oy 2222 aaccacagttaccagttaccagttaccag 2250
Db 2192 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220

RESULT 5
US-08-973-462-1
: Sequence 1, Application US/08973462B
: Patent No. 6191270
: GENERAL INFORMATION:
: APPLICANT: DROULHE, PIERRE
: APPLICANT: DAUBERSIES, PIERRE
: TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
: FILE REFERENCE: 0660-0125-0 PCT
: CURRENT APPLICATION NUMBER: US/08/973,462B
: EARLIER FILING DATE: 1998-02-06
: EARLIER APPLICATION NUMBER: PCT/FR96/00894
: EARLIER FILING DATE: 1996-06-12
: EARLIER APPLICATION NUMBER: FR 95/07007
: NUMBER OF SEQ ID NOS: 29
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 6152
: TYPE: DNA
: ORGANISM: P. falciparum

```

```

US-08-973-462-1
Query Match 2.7%; Score 132.6; DB 4; Length 6152;
Best Local Similarity 44.7%; Pred. No. 4,2e-13;
Matches 1320; Conservative 0; Mismatches 1569; Indels 66; Gaps 18;

Oy 582 tctatgtcttaatgattatgttcaaatlacccttcaactcttaaatctgcgcgaatga 641
Db 2686 tcttctgtaaacgcttgcacaacaattatcagacaactctttaaagtaattatagtg 2745
Oy 642 agacgtacttaaaaaactgtgttcggaatagaaacccatlagacaatatlaagata 701
Db 2746 tctcgaactcgggaataaagaagacagtatattaaatlgatagaagaagtaaga 2805
Oy 702 tctaggaataatcgggaattacattaaaaataaaaaaacocatagaataatataatga 761
Db 2806 tctagtcaccacaatactcgaagaacgtagaagaactcagaactggaagtgtaactctt 2865
Oy 762 attaatgagaagaatgaagaacaaatltgataaaataagaatgcacactaaagaaga 821
Db 2866 tagtaacatatagagagagatcaagaataactatctactaatgatacctatagaaga 2925
Oy 822 aaaaaaaattatatacgaagctcaatgatgtcttctatcttaataaacaattagaaga 881
Db 2926 attagaagaaccccaagaagaatgtatagtgccgtctttagaataaatacccaagtaga 2985
Oy 882 agcacataatttaataagcgtttagaanaacgtatlgacactttaaanaaatga 941
Db 2986 ggaagaagaagaatagatagatgttaattgaagaagaagaagtcgtaccactt 3045
Oy 942 ccttaaggaattactgtatagaataatgaataaatacccccgcgcgaactcttg 1001
Db 3046 aatgaagaactgtggaacagcgagaagaagaagcgcaatacattacggaatatctga 3105
Oy 1002 aatataccacaaactctctcttctgataagaacaaataacggaacacagaagaagaat 1061
Db 3106 aattttagaagaagaatagcagtagaagaatgaataatgttcagagaattttagagaat 3165
Oy 1062 aaaaaaattcccaaaactataatlaata-----tttagatgttattactgattcc 1115
Db 3166 aaacgaactgttatttaactcgtatagatagaagtagagaacagtgaaattagcgg 3225
Oy 1116 acttgattagatatactatttagagaagaataaataaataattgtatagaagtagc 1175
Db 3226 agaaagtttagaagaacaaatgaatgtataagcaattttagtgaatatctgataatgc 3285
Oy 1176 tgaacaaagaagtaact---gaaccatgaatatccaaatggagttactatcctt 1232
Db 3286 aaaaagaaatacagaagaatttatttaacaggtatgttgcgaatagaaacagttatgc 3345
Oy 1233 gtcatatacgaatatacaatgtctttaaataatgaacttaattcttctgtgatttaata 1292
Db 3346 aatccaatcagaagaagaagtgatttgaatga---aatgtgttagtgcaatttaga 3402
Oy 1293 tccattgtattacaaaagaacaaagttaaaaacatatatactcgtataagaagaataa 1352
Db 3403 taatatagaataatagaagaaggttatttaataaataatagaataattcagaatcga 3462
Oy 1353 attcataaagtgaatlaagaagaataaataaataaataaataaataatgaatcgaata 1412
Db 3463 aggtgttcaagaactcgttaactcgaacatgtgaacaaatgtatatagtgtgtgattgc 3522
Oy 1413 aaaaacttcaagaagaactcgaagcttcttaaatgataataaagaagaatagtaaat 1472
Db 3523 tccgtgatgaagaatcaactttaggaatataatagtgagcgaggggttgaagaagaat 3582
Oy 1473 acttaatgaattttagtagagaacaaatcattataatataatagatttaactaatttga 1532
Db 3583 gtttttaatttgaagaat---gtattaaagtgaaatgtatgttaattactcgtagaaga 3639
Oy 1533 aatgaatggaagaatatacattataaagttgaagaacttaacacacataacttctgc 1592

```

Dh 3640 aaltgaagatgaacccggttccaagaagatgagaaaaagaacactcttagtatatttgaaga 3699
Qy 1593 atcctatgaanaatttcaacataatcttgaaaaagtttaacaaagcttcaatatatga 1652
Db 3700 aatggaagaanaaatctgtgaatctatgaggaagaagaagaatlttaacagacaagat 3759
Qy 1653 agatttcttaagaat---atagtagtgaanaagaatlaaataattataaattt 1709
Db 3760 gatagatgcagtgagaatcccatagaataatcttcagattcctaagaagaactgaatc 3819
Qy 1710 aataagcaaaatagaanaatgagatcgaacactagtgtgaataatataaagaatgaaga 1769
Db 3820 tatlaagaataaagaanaagatgttccactagtgtgtgaagaagtccaagacactgat 3879
Qy 1770 acggttttgaaaaaaatttacttaagaacgaataaaccagatgaaanaatttga 1829
Db 3880 ggaatgaagtggttgaagaatcttgaatgaanaatatggaagaggaagtltatgaaga 3939
Qy 1830 agtatctgacatgtlaaagttacaagttccanaaagtttlaatgaacaaatgacga 1889
Db 3940 tggctgttgaat---aaatgacattactagcgaacttatgagaagaactcaagattaa 3995
Qy 1890 attaaaaaagactcaattgattttaaanaatgtagaataaacaataatatacatg--- 1945
Db 3996 atgaagttagaagcagatttaataaagataatggaanaaatlaaagaatagaanaagcat 4055
Qy 1946 ttcccaattcttcaacaagaanaataaagaacacttattatgaattgtgtgaaga 2005
Db 4056 tatcagaagatttcaagaataatagatgacaaagaatgatacattagaanaagttatgt 4115
Qy 2006 aagaatltgaataatlaaagtgltcatggtcctaaggtagaatcatcatgataatgaaga 2065
Db 4116 aagaagacatgtataacgaacgactgtgagtgaagttgtgagaatttaaaagatgcgaag 4175
Qy 2066 aaaaaaacataaacaagaaggtcaatgcgataatccggaacatcaaccggaaggagaa 2125
Db 4176 aagaacagatcgaaaaaagttactgattttaaagaatcttgaagaagatalatlaaagaag 4235
Qy 2126 taacagacacgaactacaacaacccggaacaagacagatcgtcttgaaggagatt 2185
Db 4236 taaagaataatcaagaactctgaagtg---aaattttagaagattlaaagaattaaaa 4292
Qy 2186 cagttacaagcacaagcacaagaacaaacaacgaacacacacacagttacccgttac 2245
Db 4293 ctattgaacacgattttagaagagaaaaagaataagaaaagtcatttgaanaa 4352
Qy 2246 cagaagcaaaagcagacaggtcccaaccacacacacacacacacacacacacacac 2305
Db 4353 tcaagaagaagcgtgaagaataaagaatcttgaacgagatatataaagaagttact 4412
Qy 2306 ttcccaattagatttcttgaanaattatgaatt-----tttaatacttcata 2358
Db 4413 catlaagaagttgaagaagaanaaaatttagaagaatgatacgcgaattaaagaaggttag 4472
Qy 2359 atatgtcacaaatataattgtgttccacactcaactatgaacggaagaatlatlaaaca 2418
Db 4473 aacataataagtggtgtgtgcgatalaaagtggttgaagaagtgatttgaagaag 4532
Qy 2419 tataaataacaaagagagaag--aagcaaatgaatgcattgacccctgaacttat 2476
Db 4533 tagatgatttaagaaggaatattatagacatttgaaggagatatggaatttggagata 4592
Qy 2477 tgtttta-----tatacaataacataccgttaatgtatttattgtttgagcttaa 2530
Db 4593 tggataaggaagatttagaagatgtaacaacaacacttgagaagaagttgaactcttaa 4652
Qy 2531 acaatgatttatacacaactattatgaaattatgaanaagaagtggttgaattat 2590
Db 4653 aagaatgtttattatgtcat---tagcattgatagaagaacaacatgaaacaagaanaa 4709
Qy 2591 ataaacttaagataatgacaaataaataattatagagaagacggaanaaagttacca 2650
Db 4710 aagctcaagacctaagttggaagaagtatttataaagaaggtttaaagaacacaa 4769

Qy 2651 catctgtaaaactctttcaagttcacaatgcaacacattatccattcaacacactcagga 2710
Db 4770 agaaaaaaatacaaaaaaagaagaagttgatttgaattgaagataaggaaccaaagaatg 4829
Qy 2711 aaccggaagtaagtgaanaatgatacatcacatcttcaanaatttgaataatgattaa 2770
Db 4830 aatatagtaagaagttgaataatgagaagtgagaatataagagaaggtgtgagaagaagatag 4889
Qy 2771 aattatttgaanaacataatgagtccttgggaaaaaaacaanaatatataccagaattag 2830
Db 4890 aagaagatataagaagaagataaagttgaagatataagatgaagatalagatgaagattag 4949
Qy 2831 gtaaaaaagtagtgaanaacttttgaagaagaatataaagaagtagtgaatttata 2890
Db 4950 gtaagaacaaagatgagattatagattttaaagttccaanaagaagacgatgaaag 5009
Qy 2891 atgaatcttacaanaatttgaanaacttaagaactgaatgatatataatcaatgaatgag 2950
Db 5010 ttaagcgaanaaagaanaaatttagaanaaagttgaagaagtc---gttagtgccttaa 5067
Qy 2951 aatcaaaaggaagaattagaagaagatataataatlaaaaaaacttcaagttat 3010
Db 5068 aaaaacgtagacgaagtaatgataatgttcaaaaatgtgaagaagttgataaga 5127
Qy 3011 cattgattataataataataataatgaanaattgaagaagatttgaanaagaaga 3070
Db 5128 agtatcaaaagcttgaatcaaaaaatgagttactaatgtttttaaacaatacaaga 5187
Qy 3071 -cagttgtaaatataaataatgcaaaataaanaaacttacttataaagaacattaga 3128
Db 5188 tttttttagtaagttaaaacttcgtaaaaaatataaagatttgcgcacacttat 5247
Qy 3129 atcaaatggaattcacttaataaccgaagcgtgatatcaaaaacttttctgtttct 3188
Db 5248 atcgcggtgcagcattt---gcattcatatgagttggtgttcttcaactttctt 5301
Qy 3189 taacaaaaaagaagcgtgaanaatgagaacatggaanaacacatctgaanaacaaanaat 3248
Db 5302 atttcatcatgtgtatacaaatgcttccacttacttatatacaaaagtgtgacaana 5361
Qy 3249 atattgaaacattataaagagactgttlaaatalataatgtygaatcalctccattaa 3308
Db 5362 tataaataaanaataagagagagacggttattatcatttgaattga-----tatcttaa 5415
Qy 3309 aacttgaagtgagaatcaattcaaacagaagataatttgcagtttgaanaactttaa 3368
Db 5416 gaatttaaaacattatttcaacaacatgaagaagaanaaatttagtaagaagaanaataaa 5475
Qy 3369 agtatt--aagtaattagaagaanaattaaagataatttaatttagaagaagaanaa 3426
Db 5476 tgaatagaagttacaacaacaagctgagaanaaagtaattgtcaggtatacaataaana 5535
Qy 3427 ttatcacttacaagtgtgattacatcttaattgtcgaatttaaagaagaatlaanaa 3486
Db 5536 cgaagaacaacctaagttgataaataataaagatccga--aaaagaagaagcga 5594
Qy 3487 aataaanaattatca 3501
Db 5595 aatcaaaaataanaa 5609

RESULT 6
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim I.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.

OY	562	tgatgtgctaigtatcatctgttcaaaactcttcaactcttcaaatctcgtgcgaatgaa	641
Db	2442	tgctgtcgtgaaaaacgttgcacaacaattatcatgaacaactctttaagtaatttattagctg	2501
OY	642	agaagctacttaaaaactctgttctgcgatatagaaacccaattagacaatattaaagata	701
Db	2502	tatcgaaaactgtagaataataagaacgatatataatattgagatagaaagttaaaagaaa	2561
OY	702	tgtagaanaaaatggaagattacattaaanaaaaataaaaacattagaaaatatataatga	761
Db	2562	tgtagtcaccacaatactatagaanaacgttagaagaactacagctcgtgaagtgtaactctt	2621
OY	762	attaattgagaagaagttagaanaacatctgtttaaaataaagatgtgcactcaagaaga	821
Db	2622	tagtaacatatttagagagatcacagaanaatactattactattagatattacattagagaaa	2681
OY	822	aaaaaaaattatcaccaagctcaatatgtactcttatttacaataaacaattaga	881
Db	2682	attgagaagaactccagaaaatgttataaagtgcgcgctttagaanaataccaagtga	2741
OY	882	agcacataatttaaagcgttttagaanaacglatgtgacactttaaaaaaatgaaaa	941
Db	2742	ggaanaagaagaagtaatatagatgtaatttgagaagttaaagaagcgtctgaccactt	2801
OY	942	cattaaggaattactctgtatagataagataatgtaataatccccccggtcccaattcgg	1001
Db	2802	aatagaacactgttgaaacagcagagaagaagaagcgcgaataacattacgtagaataatttga	2861
OY	1002	aaatacaccaaatctctccctgtgataagaacaagaanaatccgagaaacagaaaagaat	1061
Db	2862	aaatttgaagaaaaatgtcagatagaagaataatgaaaatgttcgcagaaatttgagaat	2921
OY	1062	aaaaagaatttgccaaaacttaaatttaa-----ctgtagtlttatttcacgtacc	1115
Db	2922	aaacgaacactgtatttaactgtatagtagtaagttagaggaacacgtagaatattagcgg	2981
OY	1116	acttgaattggaataactcatttaagaagagaaaaataaaatattgtatataatgtgcagaag	1175
Db	2982	agaaagtttagaanaacaaatgaanaatgtaagaagattttttagtgaatatatttgaataatgt	3041
OY	1176	tgaacaagaagaatccaact---gaaccacaatgataccaagtgcgttacttaccctt	1223
Db	3042	aaaaggaattacaagaaaaatttattaacaggtatcgttgcagaatagaaacagtatagt	3101
OY	1223	gtcatataacgatatllacacatgctttaaagtacattaaatcttcttgcgtatttaata	1292
Db	3102	aatccaatcagaagaagaaggttgattgtgaatga---aatbgygttagtgcatttga	3168
OY	1293	tcacatttgattatcaaaagaaacccaagtlaaaacatatatactgtaataatgtaagaanaa	1352
Db	3159	taatatagaataattggaagaagaagtttatataaataattagaaaattatccaatgcagatga	3218
OY	1353	attcatataatgaatatgaagaaaaaattaaatagaaaaaaaanaaattgtaatctgata	1412
Db	3219	aggtggttccaagaacgttcaactgacatgtagaacaacatgtcatctgtgtagtctgtagt	3278
OY	1413	aaaaacttcaagaacagacgtctagtcctttaaagtataacaacaagaatatgtaaaaaat	1472
Db	3279	tcgcgcataagaagatccaattttttagaatataaatgtgcgcagaaggttgaaagaat	3338
OY	1473	acttaatgaaattatgatatgtagcaaatccaataaatalagatlttaactaatcttgcanaa	1532
Db	3339	gttttctaatttgaagaat---gtatttaaaatggaagtgtgtagtcttattactgtaga	3395
OY	1533	aatgatgtgtaaaaagatatcatataaagtttgaagaactttcacacacataactttgc	1592
Db	3396	aattagaatgtgaacccgttccaanaaaggttgaanaaagaacgtttagatatttgaaga	3455
OY	1593	atccctatgaanaatcttcaacataatcttgaanaagtttaacaaaagctcctaataatata	1652
Db	3456	aatggaagaaaaattgttagatgtatttagaggaagaaaaagaagattcaacagaagaat	3515
OY	1653	agattactctttaagaagat---atagtaattgaaaaagaattaaaatatlaaaaaatt	1709

Dd	3516	gatagatagcagtagaagaatccatccatagaaatattcttcagattcttcaagaagaaactgtaatc	3575
Oy	1710	aataagcaaaaatagaagaatagatatgaaacattagttcttgaaaaatatcaaaaaagatgaaga	1769
Dd	3576	tatttaagaataaagaaaaaaatgcttccactagttgtcttgaaagaagttcaagacatgatata	3635
Oy	1770	acagcttttggaaaaaaaattactcaagaacgaaaaataaccagataggaaaaatlttga	1829
Dd	3636	ggatgaaagatgtcttgaaagaattcttagaattgaaaaatactgaaagagaggttaactgaaga	3695
Oy	1830	agttatcgactctgttaaaagtatacagttccaaaagtcttatataagacaaaatttgaga	1889
Dd	3696	tgctgttgtaaat---aaatgacattactgccaacttattgaagaacatccaaagttaaa	3751
Oy	1890	attaaaaaagatccatcttgaattttaaanaaagttagatttaaacataatatacatg----	1945
Dd	3752	atgaaagtgcgaacgcgattttataaagaatatagynaanaattaaagaatttagaaaaagcat	3811
Oy	1946	ttcccaattcttacaacaaagaataaagaacgaacccattatttlaatttgtytgtaaaa	2005
Dd	3812	tatcagaagatcttcaagaagaatatagatgcagaagaatgatacattggaaaaagtattg	3871
Oy	2006	aagaaattgataaatttaaagtgttcatctgcctaaagttagaatcattgtataaattgaaaga	2065
Dd	3872	aagaggaaatgataataacgacagcagttgtagaagttgttagaatlttaaaagatgtcgaag	3931
Oy	2066	aaaaaaacataaaaacgaaagcgtccatcgatbaattcggaccatccaaccgaaaggaaga	2125
Dd	3932	aagacaaatctgaaaaaagtctcgtatttaaagaaccttgaaagaatataatataaagaagag	3991
Oy	2126	taacaggacaagacactacaaaacctgycacaacaaagcagatctgtttagaagagaatt	2185
Dd	3992	taaaagaatccaagaacttgtaagtg---aaattttagaagattataaagaatlaaaaa	4048
Oy	2186	cagtaacaagcgaagcgaagaacaaaataacagacaaaccacgaatgccagtaccagtac	2245
Dd	4049	ctattgaaacacgataatttttagaagagaaaaaagaataagaaaaagatcaatttgtaaat	4108
Oy	2246	cagaaagcaaaagcgaagctcccaaccacgaacgaacgaaataataataaactgaanaag	2305
Dd	4109	tcgaagaagaagctcgagaataataaagaccttgaaacgagatatattaaagaagatattc	4168
Oy	2306	ttccaattagattactctgtaaaaaattatgatatt-----tttaatacttcatat	2358
Dd	4169	cattagaagttgaaagaagaaaaaaatttagaagagttacaagatataaagaagaagtgag	4228
Oy	2359	atagtccaaaatatatttggttttccactccaactatgaacgnaaagatatlaaaaaa	2418
Dd	4229	aacataaatatagtgtagcgcacataaaaagtggttgaaagaagaatgatbagaagag	4288
Oy	2419	tataaatattcaaaagggggagaa---aagcaaatlaagttcatgtagccattagaactat	2476
Dd	4289	tagatgattttaaagaagaatatacttagacatgtttaaaggagatatagaaattaggggata	4348
Oy	2477	tgatttaa-----tatacaaaataacataccctgtaaatgtatctatgtttagagcttaa	2530
Dd	4349	tgagtaaggaagaagtttgagaagatgtaacacaaaactctggagaagaagttgaaccttaa	4408
Oy	2531	acaaatgattatccaactattttagyaaatttatgaaaaaagaaatggttgytaattat	2590
Dd	4409	aagatgtttattatctagygcatla---gycatgagtgaagaacaaatgaaacagaagaaa	4465
Oy	2591	ataaacttaagataatagcaaaaataaaaaatttatitgagaagacgaaaaaagtatcca	2650
Dd	4466	aagctcaaaagaccatgaagttggaagaagatttataaaaagaagaaggttlaaagaagacca	4525
Oy	2651	catctgtaaaagacacttcttaagttcaatgcaaccttatcatcatcaaccctcagata	2710
Dd	4526	agaaaaaaataacaaaaaagaagaagtatggtattgatatataagaataagaaaccaaaagtg	4585
Oy	2711	aaccgaagtaagtgcaaatgatactacatccattccaatttgatatagtttaa	2770

D	4586	aatagtaggaagcttggaatggaagatggaagatagagaagatgtagaagaagatag	4645
Q	2771	aattatttgaaaacacatattgagctcttggaaaaaacaaaaatatataccaagaattag	2830
D	4646	aagaagatattagaagaagataaagcttggaagatatagatgaaagatatagataagatag	4705
Q	2831	gtcaaaaagttagtggaaaccttttagaagaagatataaagaatgatatcatcttata	2890
D	4706	gtgaagaacaaagatgagttatgatttatatgctccaaaagaagaacgcgtgaaaag	4765
Q	2891	atgacattttacaacatttvgtaaatctcaaacgtacgtatgatatcaatcatgtagtg	2950
D	4766	ttaaagcgaaaaaagaaaaaatatgaaaaaaaagcttgaagaagct--gttagtggtctaa	4823
Q	2951	aatcaaaaaggaagaattagaagaagatatataatcaatlaaaaaaaccttacaagtat	3010
D	4824	aaaacacgtatgaagaagtatagtgaatgtctcaaaaaattgataaagaagctgttaaaga	4883
Q	3011	catttgctttataataataataataataatgtgaagaagttattttagataaagaaga	3070
D	4884	agtacttaagcttctggaatcaaaaaatgagtctcaatgcttttaaaaaacaataaga	4943
Q	3071	--cagttgtgtaataataaatgtcaaatlaaaaaacttataatlaaaaaagaacattaga	3128
D	4944	tttttttagaagaagttaaaaaactctgtaaaaaatatagaagatttgcgcacattcat	5003
Q	3129	atcaaaaatgaaatcacttaataataccaagaacgtatatacaaaacttctgtttctt	3188
D	5004	atctgcgcgttgcagcatttg-----catcatatgtagtgtgtctttacattcttctt	5057
Q	3189	taacaaaaaaaagaagcctgaaatagcagaactgtaaaaacacattagaagaacacaaat	3248
D	5058	atttcactgtgtatacaaatagctcttccaacttaactattatatacaaaagcttgcaaac	5117
Q	3249	attattgaaacattataaagagctgttlaaatatataatagttgtaatcatcttcaataa	3308
D	5118	tataataaaaaataagagagagaccgtttattcatgttatttg-----tatctttaa	5171
Q	3309	aacctttagtgaagatcaatctcaaacagaagataataatagccagtttagaaaaacttaa	3368
D	5172	gaattttaaaccattatcttcaacaaaatgaaaagaaaattttagtaagaaaaataataa	5221
Q	3369	agttat--aagttaattagaaggaanaattaaaggaatttaaatctttagaagaagaaaaa	3426
D	5232	tgtaataagaaagtatacaaaccaagccttgagaaaaaagtgtaatgtacacgtgaacaaataaac	5291
Q	3427	ttatcatattatcaagtgtgattatcatcttlaattgtcgtgaattaaagaagtataaaa	3486
D	5292	cgaagaaacacacaaagttgataaaaaataataagttaccgaaaaaaaagaagaacgaaaa	5351
Q	3487	aataaaa 3493	
D	5352	atcaaaa5358	

RESULT 8
 US-08-487-826B-13/C
 : Sequence 13. Application US/08487826B
 : Patent No. 5993827
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Sim, Kim L.
 : APPLICANT: Chlunis, Chetan
 : APPLICANT: Miller, Louis H.
 : APPLICANT: Peterson, David S.
 : APPLICANT: Su, Xin-zhuan
 : APPLICANT: Wellens, Thomas E.
 : TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 : TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 : TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 : TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 : NUMBER OF SEQUENCES: 45
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Knobbe Martens Olson & Bear
 : STREET: 620 Newport Center Drive
 : CITY: Newport Beach

STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

[illegible]

[illegible]

```

RESULT      9
US-08-323-170B-1
: Sequence 1, Application US/08323170B
: Patent No. 5733772
:
: GENERAL INFORMATION:
:
: APPLICANT: Williamson, Kim C.
:
: APPLICANT: Kaslow, David C.
:
: TITLE OF INVENTION: Cloning and Expression of Plasmodium
:
: TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, Pf5230
:
: NUMBER OF SEQUENCES: 4

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323,170B
FILING DATE: 13-OCT-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/010,409
FILING DATE: 29-JAN-1993

ATTORNEY/AGENT INFORMATION:
NAME: Quine, Jonathan A.
REGISTRATION NUMBER: P-41,261
REFERENCE/DOCKET NUMBER: 015280-113100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9636 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 149..9556
US-08-323-170B-1

```

[illegible]


```
QY 1117 cttgaattgaactatttaagagaaaaaataaataatgataaagtcgaaggtt 1176
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2040 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1981
QY 1177 gaaacaaaggaatcaactgaacccaatgatatccaaatgaggttaactcttgta 1236
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1980 AATGAGAAAAAGTGGTTAACTACATGTTTCTTTTCTTTTCTTTTCTTTT 1921
QY 1237 tatacgatatacaactcttaacttaacttaactctcttggtatataatccaa 1296
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1920 ACTTTTAACTTTTAACTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1861
QY 1297 ttgtatatacaaaagaaacaaagtaaaacataatcgtatgaagaaacaaatc 1356
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1860 TTTTAAAAAAGAAAAAGAAA---AAAAAAGAAAGTACATTTTAAATTTAAATTTA 1804
QY 1357 atcaatgaattaaagaaataaataagaaaaaaataatgaaatcgtataaaaa 1416
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1803 TTCAATCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1744
QY 1417 tcttcgaagaaagatcgaactcttaactgaatgatatcaaaagaatgaataact 1476
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1743 AAAAAGTAAAGAAAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1684
QY 1477 aatgaatctatgataagcaaatcaataaataatgataatgaatgaatgaatga 1536
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1683 AAAAAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1628
QY 1537 atggttaaaagatatcatataaagttgagaacactacacacataactcttgatcc 1596
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1627 AAGAGCGGAAAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1568
QY 1597 tatgaataatcttaacataactctgaagaaagttacaaagaagccttaactatgaagat 1656
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1567 ACCTTTAACTTTGAAATCAATGATATTAATTAATTAATTAATTAATTAATTA 1508
QY 1657 taattcttaaggaataatgaatctgaagaaag--gaatcaataatataaataatga 1714
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1507 TTTTCTTTTCAACATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1448
QY 1715 gcaaaatagaaaaatgagatctgaacactgaatgaataatcaaaagaatgaagaaagc 1774
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1447 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1388
QY 1775 tttttgaaaaaaataatcttaagaaagaaataaaccgaatgaagaaatcttgaaagt 1834
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1387 TCGTACTGAAGTATGATCTTATTAATTAATTAATTAATTAATTAATTAATTA 1328
QY 1835 ctgacactgtaaaagtaacagttcaaaagtttaataatgaacaaatctgacgaattaa 1894
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1327 GGTAAATATATATCTTCTGTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1268
QY 1895 aaaagactcaattgattttaaaaaatgtagaatlaaaacataatataatggtcccaatt 1954
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1267 TGACAAAAAAGAAAAAAGAAATTTATTTCTTCAATGATGATGAAGACCTCGTT 1208
QY 1955 cttaacaaacaaagaataagcaagaaacctatattatattgtgtt 2000
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1207 TTTAATTAATTAATTAATTAATTTGTCTCGTCTGATGATGAATTAAGGTT 1162

RESULT 11
US-08-446-855A-1
; Sequence 1, Application US/08446855A
; Patent No. 5849573
; GENERAL INFORMATION:
; APPLICANT: Stewart, Thomas S
; APPLICANT: Flores, Maria V
; APPLICANT: O'Sullivan, William J
; TITLE OF INVENTION: Nucleotide sequence encoding cardamomyl
; NUMBER OF SEQUENCES: 2
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 1100 No. 5849573th Gleebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446/855A
; FILING DATE: 06-Jul-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mitchard, Leonard C
; REGISTRATION NUMBER: 29,009
; REFERENCE/DOCKET NUMBER: 47-80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic
; US-08-446-855A-1

Query Match 2.1%; Score 102; DB 2; Length 8920;
Best Local Similarity 42.7%; Pred. No. 2.9e-08;
Matches 1693; Conservative 0; Mismatches 2200; Indels 71; Gaps 20;

QY 489 ttcaaatcttaattgaatgagatgaagaataatgaatgaatgaatgaatgaatgaatga 548
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2241 TTTCAATATTTGTAATTTTGTAGTAATGATGATGATGATGATGATGATGATGATG 2300
QY 549 ttatttgaatttaataaggaacaaatlaaatgaatgaatgaatgaatgaatgaatgaatga 608
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2301 GATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2358
QY 609 acccttcaacttaaatctgtagaataatgaatgaatgaatgaatgaatgaatgaatgaatga 668
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2359 TAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2418
QY 669 atataagaacactgagacaaatlaaatgaatgaatgaatgaatgaatgaatgaatgaatga 728
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2419 TAATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2478
QY 729 aaaaaataaaaaaacatagaataataaataatgaatgaatgaatgaatgaatgaatgaatga 788
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2479 TAACGAGGTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2538
QY 789 tgataaaataagaatgacaaatgaagaagaagaagaagaagaagaagaagaagaagaagaaga 848
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2539 TACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2598
QY 849 tgattcttatttaacaaatgaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 908
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2599 TTTATTTCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2658
QY 909 aaaaactatgacactttaaataaaaaataaataaataaataaataaataaataaataaataa 968
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2659 AAGAAAGTAAAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2718
QY 969 tgaatataaaatcccccacggcgaatcttgagaataacaaataacccctccttgatata 1028
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2719 TATTAATCAAAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2778
```


Db 4926 TAACTAATTTATTTAACTTAACTTCAAGTCAGACATGATGTTCCCATTAATAA 4985
QY 3236 aaaaacaaaatattattgaaacattataaagacttgtaaatatttaattgtaaat 3295
Db 4986 TGAAGAAGAAAGATATGACCGCTTAATATATACGAATGCAAAATGAA----- 5036
QY 3296 cactccattaaaaacttaagttaaagaaatcaatcaacagagaataattatgcagtt 3355
Db 5037 -----AAAAAGTCATGTCAGAACCACTTATATATGAAGTAGTTGATGAAG 5086
QY 3356 tgaagaactttaagaatttaagaaatagaagaataatgaaggaatttaattag 3415
Db 5087 GATACCAATTTACACAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAAT 5144
QY 3416 aaaaagaaa--aaattacattacttaacagtgatctacacttaattgcgaattaa 3472
Db 5145 AAAATTAATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5204
QY 3473 aagaagtaataaaaaataaaattacagtaattctccaaagtgaaataacagatg 3532
Db 5205 ATACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5264
QY 3533 ttaacaaatgcatagaactcttaacaaaattctcccaagaagaacagatgtgcaag 3592
Db 5265 TAAAGTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5324
QY 3593 ttgtaagtgaaagtgatgcagacacattagaacaaagtaacccaagaacccgcacaa 3652
Db 5325 ACTTAA---ATTGTTATTTGAAAGGGCGGTAGCCATCATTAATTAATTAATTAATTAAT 5381
QY 3653 ctcatgtagagacagagcttaacacacataacacataacagtgatgaagag-- 3710
Db 5382 AGAGTAGAGATCTGATGATACAAATATTTTAAGTGCACAAATTAATTAATTAATTTT 5441
QY 3711 -----agatgacgtaacatagtaactatatttgagagaatcogaagaatgatg 3763
Db 5442 CATGTAAATGATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5501
QY 3764 attagagaagtagtaacagagaagcagtaactctccatccgaattgtaacatactt 3823
Db 5502 CGAAAAACGACAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5561
QY 3824 cttaaaattgaaatgaatlgaggttttaatttaaaaccttaagcaggtgtttataga 3883
Db 5562 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5621
QY 3884 gttaaaaaaacaaatagaacaaatgacatgacatlttaagttaagttaaggaattt 3943
Db 5622 TGGTAAATTTAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5681
QY 3944 taattacagatttaaaacgtgaaatcttaaaaaatttttaagaatcgaatttaattc 4003
Db 5682 AAAAGATTTGATGTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5738
QY 4004 catataaagatttaacaaatgaatltgltgcaaaagatcatalaaattcttaata 4063
Db 5739 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5798
QY 4064 aagaaaaaagagataaattcttaagcagatlaattataatgaagattcaatgaatg 4123
Db 5799 ATTCAACTAATGATCTTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 5858
QY 4124 atataaatttgcaaatgttcttgatattataaaatattatccgaaataataataat 4183
Db 5859 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5918
QY 4184 cagatttagattcaatlaaaaaatataacagacacaaaggttgaaatgagaaatc 4243
Db 5919 GTAATTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5975
QY 4244 ttcccttttaacaaatattgagaccttatataaaacagtttaattgaatgat-ctt 4302
Db 5976 TGAAGAGTTAATGATGAGAAATTTAAATGATGAATTAATTAATTAATTAATTAATTAAT 6035

QY 4303 ttgttaactttagaagcaaaagttcctaattatatacatatagagaatcacaagtagaa 4362
Db 6036 ATTGTTCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6095
QY 4363 gttaaaataaagaacttaattacttaaaaaacaattcagaacaaattgacagatttaa 4422
Db 6096 TGAAGAAATGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6155
QY 4423 aaaa 4426
Db 6156 ACA 6159

RESULT 12
US-09-150-741-1
; Sequence 1, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
; APPLICANT: Stewart et al.
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
; TITLE OF INVENTION: Synthesase II
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/150,741
; EARLIER APPLICATION NUMBER: 1998-09-10
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: A093/00617
; EARLIER FILING DATE: 1993-12-02
; EARLIER APPLICATION NUMBER: 08/446,855
; EARLIER FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 8920
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
US-09-150-741-1

Query Match 2.1%; Score 102; DB 4; Length 8920;
Best Local Similarity 42.7%; Pred. No. 2.9e-08;
Matches 1693; Conservative 0; Mismatches 2200; Indels 71; Gaps 20;
QY 489 ttcaaatatttaattgagagatgaagaatgaatgaattatataataaactt 548
Db 2241 ttcaaatattgtaattttagtaagtagagataaaatgattcttlttaagtata 2300
QY 549 ttatttgatttaagagcaaaatlaaagtgtatggtcgaattgatttgcaat 608
Db 2301 gtatagtgatattgataa--atatttaattgaccttgaagaataatgacttcaat 2358
QY 609 accttcaactttaaattctgtcaaatgtaattagacgttaactaaacttggtcg 668
Db 2359 taataatgtatgatatgatatgatatgattgttaataaaatacaatatctatctaa 2418
QY 669 atatagaaaacacatgacaaatlaaagataatgtagaagaaatggaagatttaact 728
Db 2419 taataaatatgacaaacacacataacgaataaatacaaaatacaaaatacaacaa 2478
QY 729 aaaaaataaaaaaacatgaagaataataatgaatttaattgaagaaggaagaacaaat 788
Db 2479 taacaggttgattatataaagaagaatgaggaataatgltcaatagtaaggtcttta 2538
QY 789 tgaataaaataagaatgcaactaaagaagaagaagaaaaaaatlaacaaagctcaata 848
Db 2539 tagcaatatataatgaatgacacaaataatgaacataccgaatttaatttaaatga 2598
QY 849 tgatcttctatttaacaaataaacaattagaagaacacataattataaaggtttta 908
Db 2599 ttattctacttattatagaagaagaatgaaataatgaaatcttcaatttglttaacaa 2658

QY	909	aaacgctttgacccctttaaaaaaaatgaaaacatlaaggaattactgtagaataa	968
Db	2659	aagaaaagtagacacataaagaanaaaactgtattgtgtgtgtgtgtataaaaaatag	2718
QY	969	tgaatttaaaatcccccacggccaaactctgtgaataacccaatactctctgtataa	1028
Db	2719	tataatcaaaaatttaataagacacggtatgagctcttcattacaatattatgtacc	2778
QY	1029	gaacaaaaaatcgaggagacacgaaaagaataaagaatctgcacaactatnaat	1088
Db	2779	ttattattacacatttlaactatagatagttatgtgtgagcttttattctaaatgycgcg	2838
QY	1089	taaatgtatgatttatttactgtacatccacttgtaagaatctatattgaagaaaaa	1148
Db	2839	agatccctaaaaagtgtaattccctctlaaaaaaatttgaagaatgtttacaacaataa	2898
QY	1149	taaaatatgataaagtgcaaaaggttgaacacaagaatcaactgaaaccaatgaa---	1295
Db	2899	aattatttggtaattcgtttagtgaatcaactataggatatacatggtgttggaac	2958
QY	1206	atatacaaatggaattactaactccttgatcataaagatlaaacaatgycctttaatga	1265
Db	2959	ataaaaaatgaaatagtgtaatagaggtgtgtatcaaccgtaataatgtagataa	3018
QY	1266	acttaattccttggtgatttaatttaaccatttgatttacaagaacaaatgaaaa	1335
Db	3019	tatatgtacattccccaacaaatcgtgatctgtttaagaanaaaatcattttaa	3078
QY	1326	catatatctgataatgaagaanaaaatcctaataatgtaataatggaanaaattaaat	1385
Db	3079	aagaaagaagcttgcgattgtatataaaatgctaaatgtaaatcatalagaaggaattc	3138
QY	1386	agaaaaaaaataatgtatctgataaaaaatcttcgagaagaatctaaagctttaa	1445
Db	3139	acataaaaatggaagatttatagtgtgcagcttcaactcgtgggtataaatgyccca	3198
QY	1446	tgatatacaaaagaatataatgaanaaatctacttaatgaattttgat-agcaaatcaata	1504
Db	3199	agataccatcttttactttaagaattctctttagatctcttaataagaanaaacaata	3258
QY	1505	ataatatagatttaacaaatttcgaaaaaattgtgtgaagaatattcatatgaattg	1564
Db	3259	tagagatatatttagatataatataatttatataaaaaagaagtgctctttagagag	3318
QY	1565	agaaactcacacacataacttcttgcatcccatgaataaattctaacaatacttgaa	1624
Db	3319	tgtgtgtttatgataagacaaagcagaggaattcgatttcaagaaaccaagaacataa	3378
QY	1625	agttacaagaagctcttaaatatataatggaagatattctttaagaaatagtagtgtaa	1664
Db	3379	aagtttaaaagaatgtgatatatgttatattatagt-taatcctaacaagaactgttc	3437
QY	1685	aagatttaaaatttatataaaatttataagaacaaatagaanaatgagttagaactatag	1744
Db	3438	aaacccaagaagtttgcgcgaataagtgataacttcttccagtttaattgtgaatttag	3497
QY	1745	ttgaataatataaaaaagatgaagaacagcttcttgaanaaaaaattctaagaagcgaa	1804
Db	3498	aaaaaattattataaaagaagaaacccgattttattttatgatacttggtygtcagaag	3557
QY	1805	ataaacaagatgaaaaaatttgaagatctgacatgtgtaaagtacaagtccaagaag	1864
Db	3558	ctttaaatgtgcttgaattgttagatccaataaaagtattgaaaaagaataattgtcaat	3617
QY	1865	tttatcataatgaaacaaatgtgacgaatttaaaaaagcctaattgttttaaaaaatgtag	1924
Db	3618	gtttaggtacatcttttgaatctataagaataacgaaaatagaacattatttctgtaa	3677
QY	1925	aattaaacataataatagcttcccaattcttcaacaacaagaanaatagaagaactt	1984
Db	3678	aattaaaaagaataatgaagaagaatgctccatataatgtgtatgtgtcaaaaatgtatcaag	3737
QY	1985	attatttaattgtgttgaanaaagaatattgataaatttaaaatgtgtcagtccaagttag	2044

[illegible]

Db 4806 ataataataatgctcgtgaggttacagaaaatgcttatgaaatagagaa 4865
QY 3118 gaaca--attagatacaaatlgaattcaacttatacccaagaacgtatccaact 3175
Db 4866 agctagattatccacacataaagttatgtaacttatacgcgcgaatccgcgt 4825
QY 3176 ttctgtttctttaaacaanaaagaagctgaatagcagaactgaaacacatlag 3235
Db 4926 taactaattattatatttactatcaagtcgaagacgtatgtctccattaaata 4885
QY 3236 aaaaacacaaatattatgaaacattataaggactgtttaaattataatggtgaat 3295
Db 4986 tgaagaagaaaagatgacgcttaataataacgaatgcaataaga----- 5036
QY 3296 catctcatataaacttttaagtgaaagatcattccaagaagaatataatgccaat 3355
Db 5037 -----aaaagctccgtcgaagaaccattatataagaaagttatgataag 5086
QY 3356 tagaaaactttaagatttaagtaaatlaagaagaaatlaaagaaatttaattag 3415
Db 5087 gatacacaattacacagaagaataataataataataat--atgaattcgtgaatgtag 5144
QY 3416 aaaaaga--aaattatcattatcaagtgattatacatttaattgtcgtataa 3472
Db 5145 aaataaatgttaattgtaataagaatccatgtctataataattcttcaattgtatca 5204
QY 3473 aagaagtaataaaaataaataatatacaggtatctccaagtgaataataatagcatg 3532
Db 5205 atacaataatataataataagaataattgtcaatgataatacctatacaaaaata 5264
QY 3533 ttaacaatgcatatagatcttaacaanaattctccagaaggaacagatgttcacag 3592
Db 5265 taaaagttacaataaacaattccaattatctatcgaataatgaaatggttgaacaa 5324
QY 3593 ttgtgaagtgaagtgatccgacacatttagaacaagtcacaagaacacagcatcaa 3652
Db 5325 acttaa---atgtgtatctgtaaaaggccgttagcattatataatgtaaaagaana 5381
QY 3653 ctcaatgtagagagagatctcaacaataacacacatacaaatgtcgtatgaagt-- 3710
Db 5382 agagtatgagatctgtatgatacaaatattttaagtgcacaaaattcaaaaataacttt 5441
QY 3711 -----agatgacgtatcatatgacttaatttggaagatccgaagaagattatg 3763
Db 5442 catgtaataatggaataatgaataaagaacagttgatgttaatgtaactgaanaatgata 5501
QY 3764 atttaggacagatgataacagaggaagcagtaactccctcgttaattgataacatactt 3823
Db 5502 cgaaaaaacgagagataataatactacaacagtatatttggaaggtaaaaatagtgta 5561
QY 3824 ctaaaattgaaaatgatatlgaggttttatatttaaaacctttagagaggtttatagaa 3883
Db 5562 ttaataataagaataaagaataagttcttatttgaaagtgatgaagaacatgta 5621
QY 3884 gtttaaaaaacaatlagaaaataacgtlatagacatttaagttaatglttaagatactt 3943
Db 5622 tggtaatttaaaaaagaataataataatgtaataatgtaataatgtaactgtaagga 5681
QY 3944 taaatccagatttaataaagtgaaaattccaanaatgltttagaatacagatttaattc 4003
Db 5682 aaaaagatatagga tggaaaataataataatgatatgata--aacataaagaanaata 5738
QY 4004 cataataagaatttaacataagaattatgtgtcaagaatccatataaaatttcttaata 4063
Db 5739 aatataaagatataggatataataataatagtagtggtatccaaatggaacacac 5798
QY 4064 aagaaaaaagaga taaattcttaagcagttataatataatlaaagattcaatagatcag 4123
Db 5799 attcaactaatgattcattataatttagaataatttaataacatgaaatggaatgga 5858
QY 4124 atataaatttgcaaatgagttcttgatataataataatataatccgaaaataataat 4183
Db 5859 ataataaaaatagataatgatttatactaaagaaaagataataataataaaccctg 5918

QY 4184 cagatttagattcaattraaaaaatataatcaacgcaacaacaggtgaaatagaaatacc 4243
Db 5919 gtaattcttattatgttgatattcgtatcatataataatgatacaaat--ataaga 5975
QY 4244 ttcccttttaacaatattgagaccttataataaacaagttaaatgataatgatt- 4302
Db 5976 tgaaggttaattagataaagaataatttaaatgattgataataataatgtaattga 6035
QY 4303 ttgttaattcatttagaagcaaaagtcttaattataacataatgagaatccaagtagaa 4362
Db 6036 atgttctaattataataatgtagtgcatttgtaaatggaagtagaagaatgataat 6095
QY 4363 gttaaaataaagaacttaattacttaaaaacaattccaagaacaaatggcagatttaa 4422
Db 6096 tagaaaatgattgataaaaaaataatgatacattatacaaaaacattataatcgttaa 6155
QY 4423 aaaa 4426
Db 6156 acaa 6159

RESULT 13
US-07-792-865D-1
; Sequence 1, Application US/07792865D
; Patent No. 5646247
; GENERAL INFORMATION:
; APPLICANT: John W. Barnwell, Mary W. Galinski,
; APPLICANT: Samuel P. Wertheimer
; TITLE OF INVENTION: MEROZOITE ANTIGENS LOCALIZED AT
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: Darby & Darby P.C.
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
; COMPUTER: IBM or IBM-compatible
; OPERATING SYSTEM: PC/MS-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/792, 865D
; FILING DATE: 19911004
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US90/01849
; FILING DATE: April 3, 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson, Arthur
; REGISTRATION NUMBER: 34,354
; REFERENCE/DOCKET NUMBER: 5986/14692-US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 527-7700
; TELEFAX: (212) 753-6237
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3763 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; DESCRIPTION:
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: P.vivax
; STRAIN:
; INDIVIDUAL ISOLATE:

DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
LIBRARY: Lambda gt 11 native P.vivax
LIBRARY: DNA expression library
CLONE: 5.3
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: This sequence corresponds to
OTHER INFORMATION: Figure 1A (sheets 1-4) in the
OTHER INFORMATION: Application, as filed.
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-07-792-865D-1

Query Match 2.18; Score 101.4; DB 1; Length 3763;
Best Local Similarity 43.3%; Pred. No. 3.e-08;
Matches 665; Conservative 0; Mismatches 841; Indels 29; Gaps 3;

QY 409 ttaactcaagaacccaatataccccaactcttgatttaactatgtaact 468
DB 1 TTGGATTAAGAAAAAGTTAAAGATACAAAGTTTGGATGAAAAAGAAAGCAATAGAAAA 60
QY 469 ttgtgataatacttcgttccaatatttaattgatatgagaataatgaa 528
DB 61 GCTTATGAAAAAATGGGAATTAACGCTTAAGAAATTAAGAAAAATGGATGCGAAAAAAC 120
QY 529 ttattataaattaaactttatttgattttaaagagcaaataaatgatagt 588
DB 121 ATGAAAAAAGAGTAGAAGAGCTCAAAATACAAATACAAAAGAAATTTTATGATCATGAT 180
QY 589 gctaagattattgtcaaatcttcaactcttaaatctcgtagcaaatgatacgtta 648
DB 181 GTTAATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 649 cttaaaaaacttggtagatataagaacattagacatatataagaataatgaga 708
DB 241 TATTAATAAATAAATGAGCAAAATTAACAGAAACGAATAGATATAGCAAGGTGATACA 300
QY 709 aaaaatgagaattcaattaaaaaataaaaaacattagaaaaataaataatgaat 768
DB 301 TCTAATTTTATTTATATACAGAACATACACAGTGTACAGAGTAAAGCTAATAATGAA 360
QY 769 gaagaaagtagaataaataatgataaataaataaataatgacaataagaagaataa 828
DB 361 CAATTTATTTATTTGCTTACAGAAAAAGAAAGCTGTGACACAGCCAGATATTAAC 420
QY 829 aaattatccaagctcaatatgattcttctattcaataaacaattagagaagacat 888
DB 421 GATTTGAAAGCATTAAGAGAGAGGTGATTAATAATTTACACCTAGTCAACAGAAAGT 480
QY 889 aattataagcgttttagaaaaacgtattgacatttaaaaaa-----aaatgaaac 942

DB 481 AATTCTATGAGGAAATGCCAAAAAATCTTAAGCATGAGAGATTGTAATTTTGAAAC 540
QY 943 attaaagaattacttgataatgaataaataaataaataaataaataaataaataaata 1002
DB 541 AATTCCGAACCTATAGCTTAAGAAATATVCAAAATATATCTCAAAAGCATTAAGTTTAAAG 600
QY 1003 aatacacaatactctcttgataagaacaaaaaaatcgaggaaacggaataaata 1062
DB 601 GAGAAATGCAAAAAACAACTTAATTAATAACAGATGAACTATTGCAAGAGTGGCAGCATG 660
QY 1063 aaagaattgccaaatacttaataatgataatgatttatttactgacactg-- 1120
DB 661 ATAGAAAGGCAAAAGCCACATTAAGAACATATTGACATCTTATGAGATGACCAATA 720
QY 1121 -----aattagaatactattagaagaaaaataaataatattgata 1164
DB 721 GATACGAGGTAGCAAAATTTGAACAAATTAATGCTGAATTTATGAATTAATAAGATGAA 780
QY 1165 agtgcagaagttgaacaaagaaatcaactgaacccaatgaataatccaatgagttact 1224
DB 781 ATTAAATCCTATTATTAAGTAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAG 840
QY 1225 tatcctttgataataagataatcaaatgctttaaataatgaacttaattctttgtagat 1284
DB 841 AATTCAAAAAGAGAAAGATTAATTAATGAGTCTTGCAAAAAATTAAGCTTAATGAGAA 900
QY 1285 ttaattaatccatttgattatatacaaaagaaacaaagtaaaacataatactgataatgaa 1344
DB 901 AGCAATTCGAATTAAGGTTAACTTAATTAAGAAATTAAGAAATTAAGAAATTAAGAAAC 960
QY 1345 agaaaaaatctcaataaattgaataaagaataaataaataaataaataaataaataatgaa 1404
DB 961 TACTTAATAAGATTAAGAGAGCGAGAAACAAAGCTATGTAAGATGATGATGATGATGAT 1020
QY 1405 tctgataaataatcttaagagaacagatcgaatcctttaaattgataaataaataat 1464
DB 1021 AAACATGAAACAACTATCACTAATATTTTCAAGCAATCTGAATTTTATGAGTGAAGAACT 1080
QY 1465 gaaaataacttaataaattatgataagcaatccaataataatagatttaactaat 1524
DB 1081 AATTCCTCAAAAAAATTAATTAAGCAAGAGACATATGAAAGAAATTTGAGCTCAAAAT 1140
QY 1525 ttcgaaaaaattgagtgaaagatatcaataaattgagaacttacaacataat 1584
DB 1141 TCTGAATTTAAACACAGGTGAAGGTTTCCAGAAAAATTAATTAATTAATTAATTAATTA 1200
QY 1585 actttgcatcctatgaaaaatcttaacataatcttgaagaattgaacaaagctttaaa 1644
DB 1201 CATTAATTTATGACAAACGAGAGATGAACTTAATTAATTAATTAATTAATTAATTAAT 1260
QY 1645 tataatgaagattatctttaaagaatagtagtttgaaaaagaattaaatattataa 1704
DB 1261 CTTATGAAACTTAACTAGAAAGTGAATTAATTAATTAATTAATTAATTAATTAATTA 1320
QY 1705 aattataaagcaataaataaataaataaataaataaataaataaataaataaataa 1764
DB 1321 CAGGAGAGGAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1380
QY 1765 gaagaaagcttttgaaaaaaaatcttaacaaagcaaaataaataaataaataaata 1824
DB 1381 TCAGAAATTAATCTCAGAGAAACCTTATGAGAAAGGTGAAGAGACAGCAAAATCT-- 1435
QY 1825 ttagaagttatgactttgataaagtagaagttcaaaaagtttataatgaataaatt 1884
DB 1436 ATGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1495
QY 1885 gacgaattaaaaagactcaattgatttcaaaaaa 1919
DB 1496 GACTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1530

RESULT 14

